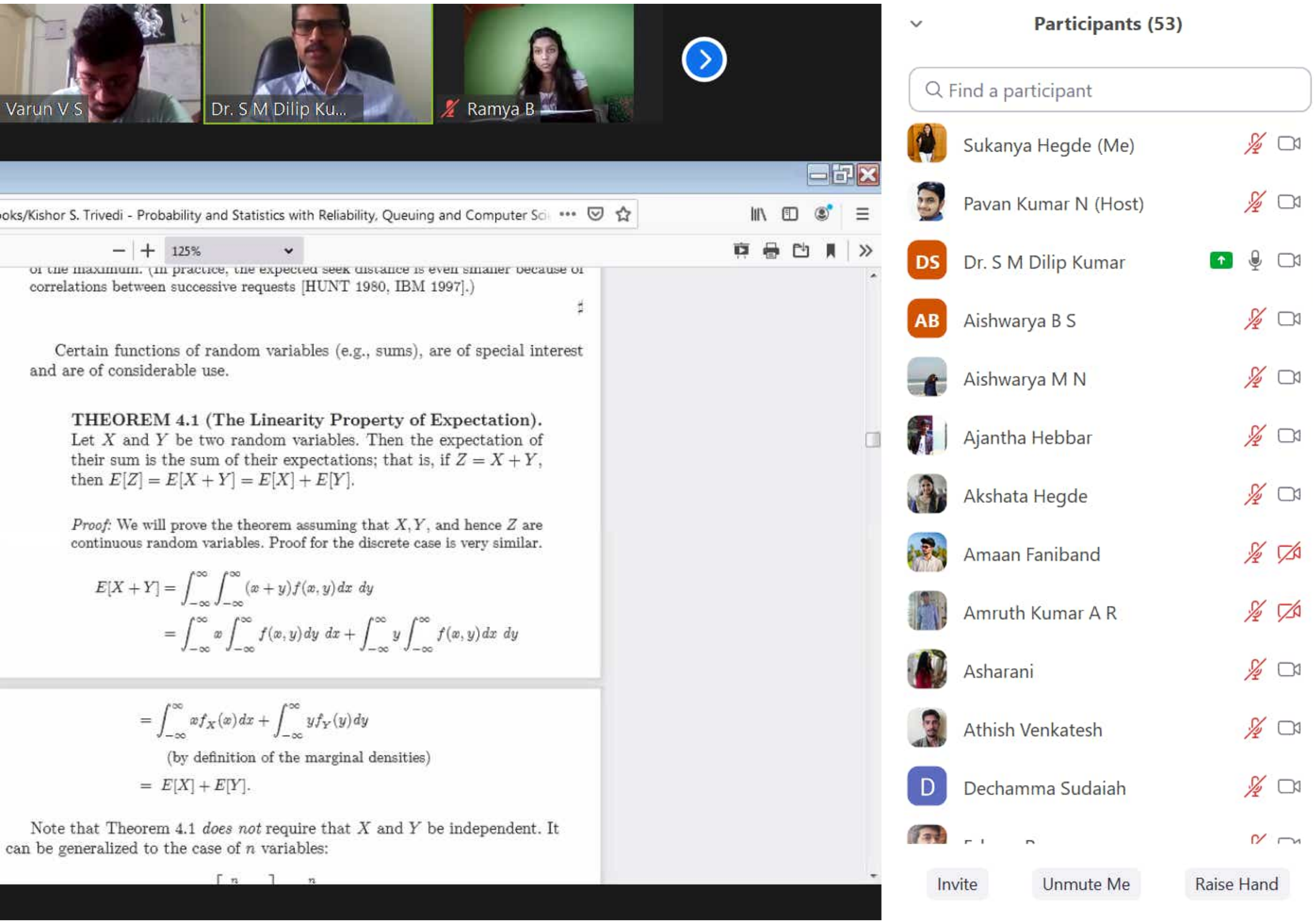


correlations between successive requests [HUNT 1980, IBM 1997].)

Certain functions of random variables (e.g., sums), are of special interest and are of considerable use.

THEOREM 4.1 (The Linearity Property of Expectation). Let X and Y be two random variables. Then the expectation of their sum is the sum of their expectations; that is, if $Z = X + Y$, then $E[Z] = E[X+Y] = E[X] + E[Y]$.

Proof: We will prove the theorem assuming that X, Y, and hence Z are continuous random variables. Proof for the discrete case is very similar.

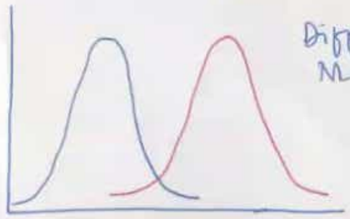
$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\,dy\,dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

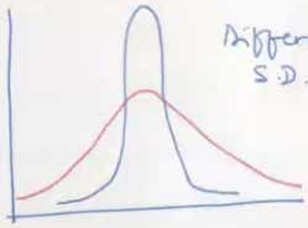
$$= \int_{-\infty}^{\infty} x f_X(x)\,dx + \int_{-\infty}^{\infty} y f_Y(y)\,dy$$
(by definition of the marginal densities)
$$= E[X] + E[Y].$$

Note that Theorem 4.1 *does not* require that X and Y be independent. It can be generalized to the case of n variables:

N.D has two parameters, the ~~shape~~ mean and S.D.

The shape changes based on parameters.





...ulation mean = μ

population S.D = σ

Sample mean = $\bar{X}$

Sample S.D = s

The standard N.D is a special case of the N.D where the mean is zero and S.D = one. (also known as z-distr).

$$z = \frac{X - \mu}{\sigma}$$

Ex:

	Apple = 110	Orange = 100
		140

population S.D $= \sigma$

Sample mean $= \bar{X}$

Sample S.D $= s$

The Standard N.D is a special case of the N.D where the mean is zero and S.D = one. (also known as z-distr).

$$z = \frac{X - \mu}{\sigma}$$

Ex:

	Apple = 110	Orange = 100
Mean Weight (gms)	100	140
S.D.	15	25

z-scores:

Apple: $\frac{(110-100)}{15} = 0.667$

Orange: $(100-140)/25 = -1.6$.

Unlike the case of expectation of a sum, the expectation of a product of two random variables does not have a simple form, unless the two random variables are independent.

THEOREM 4.2. $E[XY] = E[X]E[Y]$, if X and Y are independent random variables.

Proof: We give a proof of the theorem under the assumption that X and Y are discrete random variables. The proof for the continuous case is similar:

$$\begin{aligned} E[XY] &= \sum_i \sum_j x_i y_j p(x_i, y_j) \\ &= \sum_i \sum_j x_i y_j p_X(x_i) p_Y(y_j) \quad \text{by independence} \\ &= \sum_i x_i p_X(x_i) \sum_j y_j p_Y(y_j) = E[X]E[Y]. \end{aligned}$$

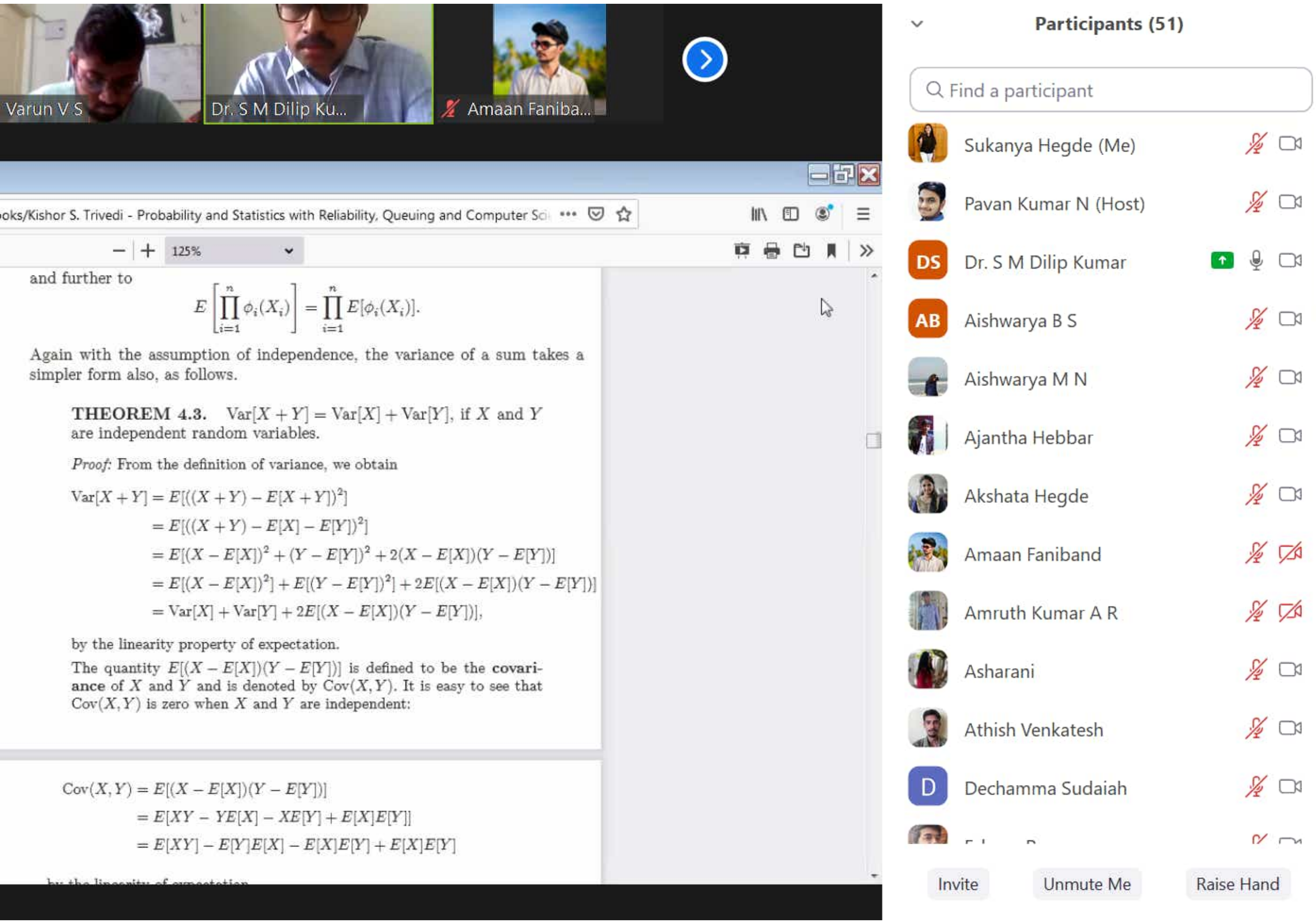
Note that converse of Theorem 4.2 does not hold; that is, random variables X and Y may satisfy the relation $E[XY] = E[X]E[Y]$ without being independent.

Theorem 4.2 can be easily generalized to a mutually independent set of n random variables $X_1, X_2, \ldots, X_n$:

$$E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E[X_i] \tag{4.14}$$

and further to

$$E\left[\prod_{i=1}^{n} \phi_i(X_i)\right] = \prod_{i=1}^{n} E[\phi_i(X_i)],$$


Varun V S
Dr. S M Dilip Ku...
Amaan Faniba...
oks/Kishor S. Trivedi - Probability and Statistics with Reliability, Queuing and Computer Sci
125%
and further to
$$E\left[\prod_{i=1}^{n} \phi_i(X_i)\right] = \prod_{i=1}^{n} E[\phi_i(X_i)].$$
Again with the assumption of independence, the variance of a sum takes a simpler form also, as follows.
THEOREM 4.3. Var$[X+Y]$ = Var$[X]$ + Var$[Y]$, if X and Y are independent random variables.
Proof: From the definition of variance, we obtain
$$\begin{aligned}\mathrm{Var}[X+Y] &= E[((X+Y)-E[X+Y])^2]\\ &= E[((X+Y)-E[X]-E[Y])^2]\\ &= E[(X-E[X])^2+(Y-E[Y])^2+2(X-E[X])(Y-E[Y])]\\ &= E[(X-E[X])^2]+E[(Y-E[Y])^2]+2E[(X-E[X])(Y-E[Y])]\\ &= \mathrm{Var}[X]+\mathrm{Var}[Y]+2E[(X-E[X])(Y-E[Y])],\end{aligned}$$
by the linearity property of expectation.
The quantity $E[(X-E[X])(Y-E[Y])]$ is defined to be the **covariance** of X and Y and is denoted by Cov(X,Y). It is easy to see that Cov(X,Y) is zero when X and Y are independent:
$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[(X-E[X])(Y-E[Y])]\\ &= E[XY-YE[X]-XE[Y]+E[X]E[Y]]\\ &= E[XY]-E[Y]E[X]-E[X]E[Y]+E[X]E[Y]\end{aligned}$$
Participants (51)
Find a participant
Sukanya Hegde (Me)
Pavan Kumar N (Host)
Dr. S M Dilip Kumar
Aishwarya B S
Aishwarya M N
Ajantha Hebbar
Akshata Hegde
Amaan Faniband
Amruth Kumar A R
Asharani
Athish Venkatesh
Dechamma Sudaiah
Invite
Unmute Me
Raise Hand

oks/Kishor S. Trivedi - Probability and Statistics with Reliability, Queuing and Computer Sci

125%

hold.[1]

Example 4.9

Let X be uniformly distributed over the interval $(-1, 1)$ and let $Y = X^2$, so that Y is completely dependent on X. Noting that for all odd values of $k > 0$, the kth moment $E[X^k] = 0$, we have

$$E[XY] = E[X^3] = 0 \quad \text{and} \quad E[X]E[Y] = 0 \cdot E[Y] = 0.$$

Therefore X and Y are uncorrelated!

♯

We have declared that $\text{Cov}(X, Y) = 0$ means X and Y are uncorrelated. On the other hand, if X and Y are linearly related—that is, $X = aY$ for some constant $a \neq 0$—then, since $E[X] = aE[Y]$, we have

$$\text{Cov}(X, Y) = a\text{Var}[Y] = \frac{1}{a}\,\text{Var}[X]$$

or

$$\text{Cov}^2(X, Y) = \text{Var}[X]\text{Var}[Y].$$

In the general case, it can be shown that

$$0 \leq \text{Cov}^2(X, Y) \leq \text{Var}[X]\text{Var}[Y] \tag{4.19}$$

using the following Cauchy–Schwarz inequality:

$$(E[XY])^2 \leq E[X^2]E[Y^2]. \tag{4.20}$$

$\text{Cov}(X, Y)$ measures the degree of linear dependence (or the degree of correlation) between the two random variables. Recalling Example 4.9, we note that the notion of covariance completely misses the quadratic dependence. It is often useful to define a measure of this dependence in a *scale-independent* success in any trial and let $q = 1\text{-}p$ denote the probability of failure in any trial. The probability of success p is constant for each trial. Then a random variable X is said to follow Binomial distribution if it takes on only non-negative values and its probability mass function is given by:

$$P(X \quad k) \quad p(k) \quad \binom{n}{k} p^k q^{n-k}; \ k \quad 0,1,2,\cdots,n \tag{7.1}$$

This is also the probability of exactly k successes in n trails in a Binomial experiment. Here p and n are known as the parameters of the distribution, in particular n is called the degree of Binomial distribution. The Binomial distribution is written as $bin(k;n,p)$.

$\binom{n}{k}$ counts the number of outcomes that include exactly k successes and $n \quad k$ failures. If q^n denotes the probability of no successes, then $1 \quad q^n$ denotes the probability of one or more successes. The probability of getting at least k successes is:

This is also the probability of exactly k successes in n trails in a Binomial experiment. Here p and n are known as the parameters of the distribution, in particular n is called the degree of Binomial distribution. The Binomial distribution is written as $bin(k;n,p)$.

$\binom{n}{k}$ counts the number of outcomes that include exactly k successes and $n \quad k$ failures. If q^n denotes the probability of no successes, then $1 \quad q^n$ denotes the probability of one or more successes. The probability of getting at least k successes is:

$$p(k)+p(k+1)+\cdots+p(n).$$

Note: $p(k) \geq 0$ for all k and $\sum p(k) \quad (q+p)^n \quad 1$. Also

$$\sum_{x=0}^{n} p(x) \quad \sum_{x=0}^{n} {}^{n}C_k p^k q^{n-k} \quad (q+p)^n \quad 1 \tag{7.2}$$

Let X be a random variable denoting the number of successes in a Binomial experiment. Then the distribution is as follows:

Table 7.2: Distribution table of X

x	0	1	2	$\cdots$	n

7.2.4 Mean and variance of the binomial distribution

Any random variable with a binomial distribution X with parameters n and p is a sum of n independent Bernoulli random variables in which the probability of success is p.

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

The mean and variance of each X_i can easily be calculated as:

$$E(X_i) \quad p.$$

$$Var(X_i) \quad p(1 \quad p) \quad nq.$$

Hence, the mean and variance of X are given by (remember that X_i are independent)

$$\mu \quad E(X) \quad np. \tag{7.3}$$

$$\sigma^2 \quad Var(X) \quad np(1 \quad p) \quad npq. \tag{7.4}$$

7.2.5 Exercises on binomial distribution

1. A biased coin is tossed 6 times. The probability of getting heads on any toss is 0.3. Let X denote the number of heads that come up. Calculate:
   (a) $P(X \quad 2)$

Hence, the mean and variance of X are given by (remember that X_i are independent)

$$\mu \quad E(X) \quad np. \tag{7.3}$$

$$\sigma^2 \quad Var(X) \quad np(1 \quad p) \quad npq. \tag{7.4}$$

7.2.5 Exercises on binomial distribution

1. A biased coin is tossed 6 times. The probability of getting heads on any toss is 0.3. Let X denote the number of heads that come up. Calculate:
   (a) $P(X \quad 2)$

(b) $P(X\ \ 3)$
(c) $P(1 < X \leq 5)$.

Solution: If we call heads a *success*, then this X has a binomial distribution with parameters $n\ \ 6$ and $p\ \ 0.3$.

$$P(X\ \ 2)\ \ \binom{6}{2}(0.3)^2(0.7)^4\ \ 0.324135.$$

$$P(X\ \ 3)\ \ \binom{6}{3}(0.3)^3(0.7)^3\ \ 0.18522.$$

$$P(1 < X \leq 5)\ \ P(X\ \ 2)+P(X\ \ 3)+P(X\ \ 4)+P(X\ \ 5)\ \ 0.324+0.185+0.059+0.01\ \ 0.578.$$

2. The probability that sachin hits a target at any time is $p\ \ \frac{1}{4}$. Suppose he fires at the target 7 times. Find the probability that he hits the target
   (a) Exactly 3 times
   (b) At least 1 time
   (c) Find expectation, variance and standard deviation.

Solution: Here, $n\ \ 7$; $p\ \ \frac{1}{4}$; $q\ \ 1\ \ \frac{1}{4}\ \ \frac{3}{4}$.

- The number of telephone calls in a week.
- The number of accidents at an intersection.

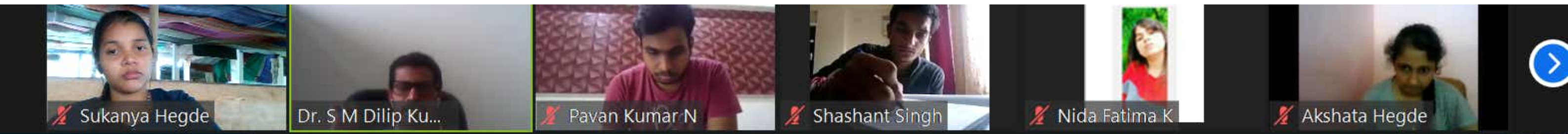
The Poisson distribution can be applied to systems with a large number of possible ev each of which are rare. For those situations in which n is large and p is very small Poisson distribution can be used to approximate the binomial distribution. The larger t and the smaller the p, the better is the approximation. If X is the random variable denc the "number of occurrences in a given interval", for which the average rate of occurre is λ then, according to the Poisson model, the probability of x occurrences in that int is given by:

$$P(X \quad x) \quad \frac{\lambda^x e^{-\lambda}}{x!} \quad x \quad 0, 1, \cdots$$

where

$P(X \quad x)$ or $f(x;\lambda$ = probability of X successes given the parameters n and p
n = sample size
p = probability of success
e = mathematical constant approximated by 2.71828
x = number of successes in the sample $(X \quad 0, 1, \cdots, n)$
$\lambda \quad np$ is called as the *rate*, *shape* or *intensity* parameter.

The Poisson random variable theoretically ranges from 0 to ∞. However, when use an approximation to the binomial distribution, the Poisson random variable - the nur of successes out of n observations - cannot be greater than the sample size n. With l n and small p, Equation 7.8 implies that the probability of observing a large numb successes becomes small and approaches zero quite rapidly. [1]



$$\sigma^2 \quad \lambda \quad np \tag{7.10}$$

Note that the variance given by Equation 7.10 agrees with that given for the binomial model (Equation 7.3) when p is close to zero so that $(1 \quad p)$ is close to one.

7.4.2 Cumulative Poisson distribution

A cumulative Poisson probability refers to the probability that the Poisson random variable is greater than some specified lower limit and less than some specified upper limit. The Poisson cumulative distribution function is given by:

$$F(x;\lambda) \quad \sum_{i=0}^{x} \frac{\lambda^i e^{-\lambda}}{i!} \tag{7.11}$$

7.4.3 Exercises on Poisson distribution

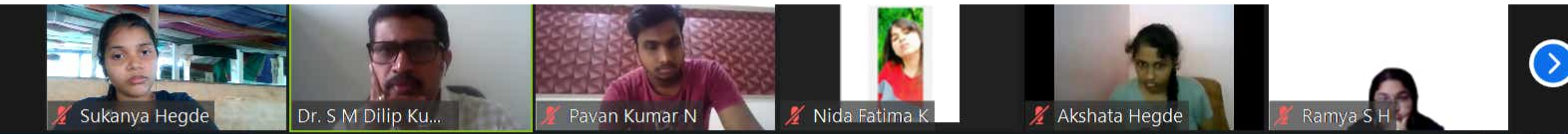
1. Consider a telephone operator who, on the average, handles five calls every 3 minutes. What is the probability that there will be
   (a) No calls in the next minute and
   (b) At least two calls in the next minute?

   Solution: Let X = number of calls in a minute, then X has a Poisson distribution with mean $\lambda \quad \frac{5}{3}$. So

   (a) The probability of no calls in the next minute = $P(X \quad 0) \quad \frac{e^{\quad \frac{5}{3}}(\frac{5}{3})^0}{0!} \quad e^{-\frac{5}{3}}$ 0.189.

   (b) The probability of at least two calls in the next minute the operator receives =

   $P(X > 2) \quad 1 \quad P(X \quad 0) \quad P(X \quad 1) \quad 1 \quad 0.189 \quad \frac{e^{\quad \frac{5}{3}}(\frac{5}{3})^1}{\ldots} \quad 0.496.$



7.4.5 A numerical comparison of the Poisson and binomial distributions

1. Ten percent of the tool produced in a certain manufacturing process turn out to be defective. Find the probability that in a sample of 10 tools chosen at random exactly two will be defective, by using
   (a) Binomial distribution
   (b) Poisson approximation to the binomial distribution.

   Solution:
   (a) The probability of a defective tool is $p \quad 0.1$. Let X be a random variable denoting the number of defective tool out of 10 chosen. Then according to the Binomial distribution
   $$P(X \quad 2) \quad {}^{10}C_2(0.1)^2(0.9)^8 \quad 0.1937.$$
   (b) We have $\lambda \quad np \quad 10 \times 0.1 \quad 1$. Then according to the Poisson distribution
   $$P(X \quad 2) \quad \frac{(1)^2 e^{-\lambda}}{2!} \quad 0.1839.$$
   In general, the approximation is good if $p \leq 0.1$ and $\lambda \quad np \leq 5$.

2. A manufacturer produces IC chips, 1 percent of which are defective. Find the probability that in a box containing 100 chips, no defectives are found.
   (a) Using Binomial approximation
   (b) Using Poisson approximation

   Solution:
   (a) Using Binomial approximation:
   Since $n \quad 100$ and $p \quad 0.01$, the required probability is:
   $$b(0;100,0.01) \quad \binom{100}{0} \times 0.01^0 \times 0.99^{100} \quad 0.99^{100} \quad 0.366.$$

5. The probability of success at each trial is same.

7.5.2 Geometric distribution

A random variable X is said to follow geometric distribution if it assumes only non-negative values and its *pmf* is given by:

$$P(X=x)=pq^{x-1};\ x=1,2,\ldots \tag{7.12}$$

In other words, suppose that independent trials, each having a probability p, $0<p<1$, of being a success, are performed until a success occurs. If we let X equal to the number of trials required, then

$$P(X=x)=p(1-p)^{x-1}=pq^{x-1};\ x=1,\ 2,\ \cdots \tag{7.13}$$

7.5.3 Mean and variance of geometric distribution

If X is a geometric random variable with parameter p, then the mean is given by

$$\mu=E(X)=\frac{1}{p} \tag{7.14}$$

and the variance of geometric distribution is

$$\sigma^2=Var(X)=\frac{1-p}{p^2} \tag{7.15}$$

values and its *pmf* is given by:

$$P(X = x) = pq^{x-1}; \ \ x = 1, 2, \ldots \tag{7.12}$$

In other words, suppose that independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs. If we let X equal to the number of trials required, then

$$P(X = x) = p(1-p)^{x-1} = pq^{x-1}; \ \ x = 1,\ 2,\ \cdots \tag{7.13}$$

7.5.3 Mean and variance of geometric distribution

If X is a geometric random variable with parameter p, then the mean is given by

$$\mu = E(X) = \frac{1}{p} \tag{7.14}$$

and the variance of geometric distribution is

$$\sigma^2 = Var(X) = \frac{1-p}{p^2} \tag{7.15}$$

7.5.4 Exercises on geometric distribution:

1. If the probability that a target is destroyed on any shot is 0.5, what is the probability that it would be destroyed on the 6^{th} attempt?
   Solution: Given $p = 0.5$; $q = 1 - 0.5 = 0.5$

A random variable X is said to follow geometric distribution if it assumes only non-negative values and its *pmf* is given by:

$$P(X = x) = pq^{x-1}; \ x = 1, 2, \ldots \tag{7.12}$$

In other words, suppose that independent trials, each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs. If we let X equal to the number of trials required, then

$$P(X = x) = p(1-p)^{x-1} = pq^{x-1}; \ x = 1, \ 2, \ \cdots \tag{7.13}$$

7.5.3 Mean and variance of geometric distribution

If X is a geometric random variable with parameter p, then the mean is given by

$$\mu = E(X) = \frac{1}{p} \tag{7.14}$$

and the variance of geometric distribution is

$$\sigma^2 = Var(X) = \frac{1-p}{p^2} \tag{7.15}$$

7.5.4 Exercises on geometric distribution:

1. If the probability that a target is destroyed on any shot is 0.5, what is the probability that it would be destroyed on the 6^{th} attempt?

$$\sigma^2 = Var(X) = \frac{1-p}{p^2} \tag{7.15}$$

7.5.4 Exercises on geometric distribution:

1. If the probability that a target is destroyed on any shot is 0.5, what is the probability that it would be destroyed on the 6^{th} attempt?
   Solution: Given $p = 0.5;\ q = 1 - 0.5 = 0.5$
   $\therefore P(X = x) = P(X = 6) = q^5 p = (0.5)^5 \times (0.5) = 0.015625.$
2. Suppose that a trainee solder shoots a target in an independent fashion. If the probability that the target is shot is 0.8.
   (a) What is the probability that the target would be hit on the 6^{th} attempt?
   (b) What is the probability that it takes him less than 5 shots?

   (c) What is the probability that it takes him an even number of shots?

(c) What is the probability that it takes him an even number of shots?

Solution: Given $p = 0.8$; $q = 1 - p = 1 - 0.8 = 0.2$

(a) Probability that the target would be hit on the 6^{th} attempt $= P(X = 6) = (0.2)^2(0.8) =$.

(b) Probability that it takes him less than 5 shots $= P(X < 5) = \sum_{k=1}^{4} q^{k-1} p = \sum_{k=1}^{4} (0.2)^{k-1} 0.8 = 0.9984$.

(c) The probability that it takes him an even number of shots is

$$= P(X = 2) + P(X = 4) + P(X = 6) + \cdots$$
$$= (0.2)(0.8)[1 + (0.2)^2 + (0.2)^4 + \cdots$$
$$= (0.2)(0.8)[1 + (0.04) + (0.04)^2 + \cdots$$
$$= (0.2)(0.8)[1 - 0.04]^{-1}$$
$$= (0.2)(0.8)[0.96] = 0.1536.$$

3. An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is the probability that:

(a) Exactly n draws are needed and

(c) What is the probability that it takes him an even number of shots?

Solution: Given $p = 0.8$; $q = 1 - p = 1 - 0.8 = 0.2$

(a) Probability that the target would be hit on the 6^{th} attempt $= P(X = 6) = (0.2)^2(0.8) =$.

(b) Probability that it takes him less than 5 shots = $P(X < 5)$ = $\sum_{k=1}^{4} q^{k-1}p$ = $\sum_{k=1}^{4}(0.2)^{k-1}0.8 = 0.9984$.

(c) The probability that it takes him an even number of shots is
$= P(X = 2) + P(X = 4) + P(X = 6) + \cdots$
$= (0.2)(0.8)[1 + (0.2)^2 + (0.2)^4 + \cdots$
$= (0.2)(0.8)[1 + (0.04) + (0.04)^2 + \cdots$
$= (0.2)(0.8)[1 - 0.04]^{-1}$
$= (0.2)(0.8)[0.96] = 0.1536$.

3. An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is the probability that:
   (a) Exactly n draws are needed and
   (b) At least k draws are needed?

Solution:

7.3.2 Hypergeometric probability distribution

Let us proceed as follows:

The x successes (defectives) can be chosen in $\binom{m}{x}$ ways.

The $n-x$ failures (non-defectives) can be chosen in $\binom{N-m}{n-x}$ ways.

The n objects can be chosen from a set of N objects in $\binom{N}{n}$ ways.

If we consider all the possibilities as equally likely, it follows that for sampling *without* replacement the probability of getting x successes in n trials is

$$h(x;\ n,\ m,\ N) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \qquad x = 0, 1, \ldots, n;\ x \le m;\ n-x \le N-m. \quad (7.5)$$

tendency of a random variable X.

Definition (Expectation). The expectation, $E[X]$, of a random variable X is defined by

$$E[X] = \begin{cases} \sum_i x_i p(x_i), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f(x)\, dx, & \text{if } X \text{ is continuous,} \end{cases} \tag{4.1}$$

provided the relevant sum or integral is absolutely convergent; that is, $\sum_i |x_i| p(x_i) < \infty$ and $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$. If the right-hand side in (4.1) is not absolutely convergent, then $E[X]$ does not exist. Most common random variables have finite expectation; however, problem 1 at the end of this section provides an example of a random variable whose expectation does not exist. Definition (4.1) can be extended to the case of mixed random variables through the use of Riemann–Stieltjes integral. Alternatively, the formula given in problem 2 at the end of this section can be used in the general case.

Example 4.1

Consider the problem of searching for a specific name in a table of names. A simple

7.3 Hypergeometric distribution

Assume we are drawing cards from a deck of well-shuffled cards *with* replacement, one card per each draw. We do this 5 times and record whether the outcome is *diamond* or not. Then this is a *binomial experiment*. If we do the same thing *without* replacement, then it is *NO LONGER* a binomial experiment. However, if we are drawing from 100 decks of cards without replacement and record only the first 5 outcomes, then it is approximately a binomial experiment. The exact model for drawing cards without replacement is **hypergeometric distribution.**

Consider the following statistical experiment. You have an urn of 10 marbles - 5 red and 5 green. You randomly select 2 marbles without replacement and count the number of red marbles you have selected. This would be a hypergeometric experiment and *not* be a binomial experiment. A binomial experiment requires that the probability of success be constant on every trial. With the above experiment, the probability of a success changes on every trial. In the beginning, the probability of selecting a red marble is $\frac{5}{10}$. If you select a red marble on the first trial, the probability of selecting a red marble on the second trial is $\frac{4}{9}$. And if you select a green marble on the first trial, the probability of selecting a red marble on the second trial is $\frac{5}{9}$. Note that if you select the marbles *with* replacement, the probability of success would not change. It would be $\frac{5}{10}$ on every trial. Then, this would be a binomial experiment.

Suppose that we are interested in the number of defective items in a sample of n units drawn without replacement from a lot containing N units, of which m are defective. Let the sample of n units be drawn in such a way that at each successive drawing, whatever units left in the remaining lot have the same chance of being selected. In other words, the probability that the first drawing will yield a defective unit is $\frac{m}{N}$, but for the second drawing it is $\frac{m-1}{N-1}$ or $\frac{m}{N-1}$, depending on whether or not the first unit drawn was defective respectively. Thus the trials are not independent.

7.3.1 Hypergeometric probability experiment

7.7.1 Probability Density Function

The probability density function (PDF) of an exponential distribution is

$$f_X(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{7.20}$$

The parameter λ is called the *rate* parameter. It is inverse of the expected duration (μ). For example, if the expected duration is 5 minutes, then the rate parameter value is 0.2. In other words, $\lambda = \frac{1}{\mu}$ and $\mu = \frac{1}{\lambda}$

7.7.2 Cumulative Distribution Function

The cumulative distribution function (CDF) of an exponential distribution is

$$F_X(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{7.21}$$

The CDF can also be written as the probability of the lifetime being less than some value x.

$$P(X \leq x) = 1 - e^{-\lambda x}$$

For instance, if $\lambda = 0.2$ and $x = 10$, then $P(x \leq 10) = 1 - e^{-2}$.

7.7.3 Mean and Variance of Exponential Distribution

The expected value of an exponential random variable is

$$E[X] = \frac{1}{\lambda} = \mu \tag{7.22}$$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{7.25}$$

where X be a continuous random variable supporting a set of positive real numbers in the range $[0, \infty)$ with parameter λ.
Note that $\int f_X(t)dt = \int_0^\infty \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_0^\infty = 0 - (-1) = 1$, which verifies that the total probability under the density curve is 1.
We generally denote the symbol μ to denote the average time needed for an occurrence. So, if λ is the specified average number of occurrences per unit time, then $\mu = \frac{1}{\lambda}$ is the average time needed for an occurrence, and $\lambda = \frac{1}{\mu}$ is again the average number of occurrences. Either parameter could be given to describe a problem.

7.7.7 Mean and variance of exponential distribution

The mean and variance of exponential distribution are respectively given by:

$$E(X) = \int_0^\infty t f(t) dt = \frac{1}{\lambda} \tag{7.26}$$

$$Var(X) = \int_0^\infty t^2 f(t) dt - E(X)^2 = \frac{1}{\lambda^2} \tag{7.27}$$

7.7.8 Exercises on exponential distribution

1. Assume that the length of a phone call in minutes is an exponential random variable X with parameter $\lambda = \frac{1}{10}$. If a customer arrives at a phone booth just before you arrive, find the probability that you have to wait
   (a) less than 5 minutes
   (b) greater than 10 minutes
   (c) between 5 and 10 minutes.

   Also compute the expectation and variance.

   Solution:

$f(x) = \frac{1}{120}e^{-\frac{x}{120}}$, $x \geq 0$.

(a) $P(X < 24) = 1 - P(X \geq 24) = 1 - \int_{24}^{\infty} f(x)dx = 1 - \int_{24}^{\infty} \frac{1}{120}e^{-\frac{x}{120}}dx = 1 - \frac{1}{120}\left(-120e^{-\frac{x}{120}}\right)_{24}^{\infty} = 1 - e^{-0.2} = 0.1813.$

(b) $P(X \geq 180) = \int_{180}^{\infty} \frac{1}{120}e^{-\frac{x}{120}}dx = \frac{1}{120}\left(-120e^{-\frac{x}{120}}\right)_{180}^{\infty} = e^{-\frac{3}{2}} = 0.2231.$

7. Calls to a customer service line arrive according to a Poisson distribution with an average of 6 calls per minute. From any initial point of observation, let X denote the time until the first call.

(a) Give the distribution of X, its *pdf* and *CDF*?

(b) What is the probability that the first call arrives

i. within 15 seconds

ii. between 6 to 12 seconds from the initial observation

iii. at least 10 seconds after the initial observation

(c) What is the average and standard deviation of the time needed for the first call?

(d) What amount of time t is such that the first call arrives within $t \leq 90\%$ of the time?

Solution:

7.8.1 Probability density function

Let $a < b$ be two real numbers. The continuous random variable X is said to have a uniform distribution on $[a,b]$ if and only if

$$f_X(x) = \begin{cases} \frac{1}{b-1} & a \le x \le b \\ 0 & x > b \end{cases} \tag{7.28}$$

Thus, the uniform distribution has two parameters a and b that define the lower and upper limit respectively of the range of X. All allowable values of the random variable, i.e. all points in a,b are equally likely or probable. The uniform distribution is also known as rectangular distribution, since the curve $f(x)$ describes the rectangle over the x-axis between the coordinates $x = a$ and $x = b$. For the above example, the uniform distribution on $[20,30]$ is

$$f_X(x) = \begin{cases} \frac{1}{10} & 20 \le x \le 30 \\ 0 & x > 30 \end{cases}$$

and $P(X \le 21.5) = \int_{-\infty}^{20} 0dx + \int_{20}^{21.5} \frac{1}{10}dx = 0 + \frac{x}{10}\Big|_{20}^{21.5} = \frac{21.5-20}{10} = 0.15 = 15\%$.

7.8.2 Cumulative distribution function

The distribution function $F(x)$ of a cumulative distribution function is given by:

$$F_X(x) = P(X \le x) = \begin{cases} 0 & -\infty < x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & b < x < \infty \end{cases} \tag{7.29}$$

Since $F(x) = \int_a^x f_X(x)dt = \int_a^x \frac{1}{b-a}dt = \frac{t}{b-a}\Big|_a^x = \frac{x-a}{b-a}$.

Reconsider the above example of uniform distribution on $[20,30]$. Then

For $x < 20$**:** $F_X(x) = \int_{-\infty}^{x} 0dt = 0$.

For $20 \le x < 30$**:** $F_X(x) = \int_{-\infty}^{20} 0dt + \int_{20}^{x} \frac{1}{10}dt = \frac{x-20}{10}$.

as rectangular distribution, since the curve $f(x)$ describes the rectangle over the x-axis between the coordinates $x = a$ and $x = b$. For the above example, the uniform distribution on $[20, 30]$ is

$$f_X(x) = \begin{cases} \frac{1}{10} & 20 \le x \le 30 \\ 0 & x > 30 \end{cases}$$

and $P(X \le 21.5) = \int_{-\infty}^{20} 0dx + \int_{20}^{21.5} \frac{1}{10} dx = 0 + \frac{x}{10}\Big|_{20}^{21.5} = \frac{21.5-20}{10} = 0.15 = 15\%$.

7.8.2 Cumulative distribution function

The distribution function $F(x)$ of a cumulative distribution function is given by:

$$F_X(x) = P(X \le x) = \begin{cases} 0 & -\infty < x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & b < x < \infty \end{cases} \tag{7.29}$$

Since $F(x) = \int_a^x f_X(x)dt = \int_a^x \frac{1}{b-a} dt = \frac{t}{b-a}\Big|_a^x = \frac{x-a}{b-a}$.

Reconsider the above example of uniform distribution on $[20, 30]$. Then

For $x < 20$**:** $F_X(x) = \int_{-\infty}^{x} 0dt = 0$.

For $20 \le x < 30$**:** $F_X(x) = \int_{-\infty}^{20} 0dt + \int_{20}^{x} \frac{1}{10} dt = \frac{x-20}{10}$.

For $x \ge 30$**:** $F_X(x) = \int_{-\infty}^{20} 0dt + \int_{20}^{30} \frac{1}{10} dt + \int_{30}^{x} 0dt = 1$.

Together: $F_X(x) = \begin{cases} 0 & x < 20 \\ \frac{x-20}{10} & 20 \le x \le 30 \\ 1 & x \ge 30 \end{cases}$ and $f_X(x) = F_X'(x) = \begin{cases} 0 & x < 20 \\ \frac{1}{10} & 20 \le x \le 30 \\ 1 & x \ge 30 \end{cases}$.

$fx(x)$ is not continuous at $x = 20$ and $x = 30$.

$$Fx(x) = \begin{cases} 0 & x < 20 \\ \frac{x-20}{10} & 20 \leq x \leq 30 \\ 1 & x \geq 30 \end{cases}.$$

Hence, pdf is the derivative of CDF, i.e. $f_X(x) = F'_X(x)$ and CDF is the integral of pdf, i.e. $F_X(x) = \int_{-\infty}^{x} f_x(t)dt$.

7.8.4 Mean and variance of uniform distribution

Using the definitions of expectation and variance leads to the following derivations. The mean of uniform distribution is:

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{2(b-a)}\left[x^2\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)}{2}. \quad (7.30)$$

Using the formula for variance, we may write:

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \int_a^b x^2\frac{1}{b-a}dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{3(b-a)}\left[x^3\right]_a^b - \frac{(b+a)^2}{2} = \frac{b^3-a^3}{3(b-a)} - \Big(\frac{b}{\ } \quad (7.31)$$

7.8.5 Exercises on uniform distribution

1. Arrivals of customers at a certain checkout counter follow a Poisson distribution. It is known that, during a given 30-minute period, one customer arrived at the counter. Find the probability that the customer arrived during the last 5 minutes of the 30-minute period.

   Solution: The actual time of arrival follows a uniform distribution over the interval of $(0, 30)$. If X denotes the arrival time, then
   $P(25 \leq X \leq 30) = \int_{25}^{30} = \frac{1}{30}dx = \frac{30-25}{30} = \frac{1}{6}$.

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{2(b-a)}\left[x^2\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)}{2}. \quad (7.30)$$

Using the formula for variance, we may write:

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \int_a^b x^2\frac{1}{b-a}dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{3(b-a)}\left[x^3\right]_a^b - \frac{(b+a)^2}{2} = \frac{b^3-a^3}{3(b-a)} - \left(\frac{b}{}\right. \quad (7.31)$$

7.8.5 Exercises on uniform distribution

1. Arrivals of customers at a certain checkout counter follow a Poisson distribution. It is known that, during a given 30-minute period, one customer arrived at the counter. Find the probability that the customer arrived during the last 5 minutes of the 30-minute period.
   Solution: The actual time of arrival follows a uniform distribution over the interval of $(0,30)$. If X denotes the arrival time, then
   $P(25 \le X \le 30) = \int_{25}^{30} = \frac{1}{30}dx = \frac{30-25}{30} = \frac{1}{6}.$
2. Electric trains on a certain line run every half an hour between mid-night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least twenty minutes?
   Solution: Let the random variable X denote the waiting time in minutes for the next train. Given that a man arrives at the station at random, X is uniform distributed on $(0,30)$ with density
   $f(x) = \begin{cases} \frac{1}{30} & 0 < x < 30 \\ 0 & x \ge 30 \end{cases}$
   Thus the probability that he has to wait for at least 20 minutes is
   $P(X \ge 20) = \int_{20}^{30} f(x)dx = \int_{20}^{30} \frac{1}{30}dx = \frac{1}{30}\int_{20}^{30} dx = \frac{1}{30}[x]_{20}^{30} = \frac{1}{30}[30-20] = \frac{10}{30} = \frac{1}{3}.$
3. The time (in minutes) passenger must wait for a computer plane in a main railway station is uniformly distributed on the interval [0,60]. What is the probability that a

$$f(x) = \begin{cases} \frac{1}{30} & 0 < x < 30 \\ 0 & x \geq 30 \end{cases}$$

Thus the probability that he has to wait for at least 20 minutes is

$P(X \geq 20) = \int_{20}^{30} f(x)dx = \int_{20}^{30} \frac{1}{30} dx = \frac{1}{30} \int_{20}^{30} dx = \frac{1}{30}[x]_{20}^{30} = \frac{1}{30}[30-20] = \frac{10}{30} = \frac{1}{3}$.

3. The time (in minutes) passenger must wait for a computer plane in a main railway station is uniformly distributed on the interval [0,60]. What is the probability that a passenger waits
   (a) Less than 20 minutes.

   (b) More than 40 minutes.

   Solution: Since $20, 40 \in [0, 60]$, therefore the cumulative distribution function, with $a = 0$ and $b = 60$ is:

   $F(x) = \frac{x-0}{60-0} = \frac{x}{60}$.

   (a) $P(X \leq 20) = F(20) = \frac{20}{60} = \frac{1}{3}$.

   (b) $P(X > 40) = 1 - P(x \leq 40) = 1 - \frac{40}{60} = \frac{1}{3}$.

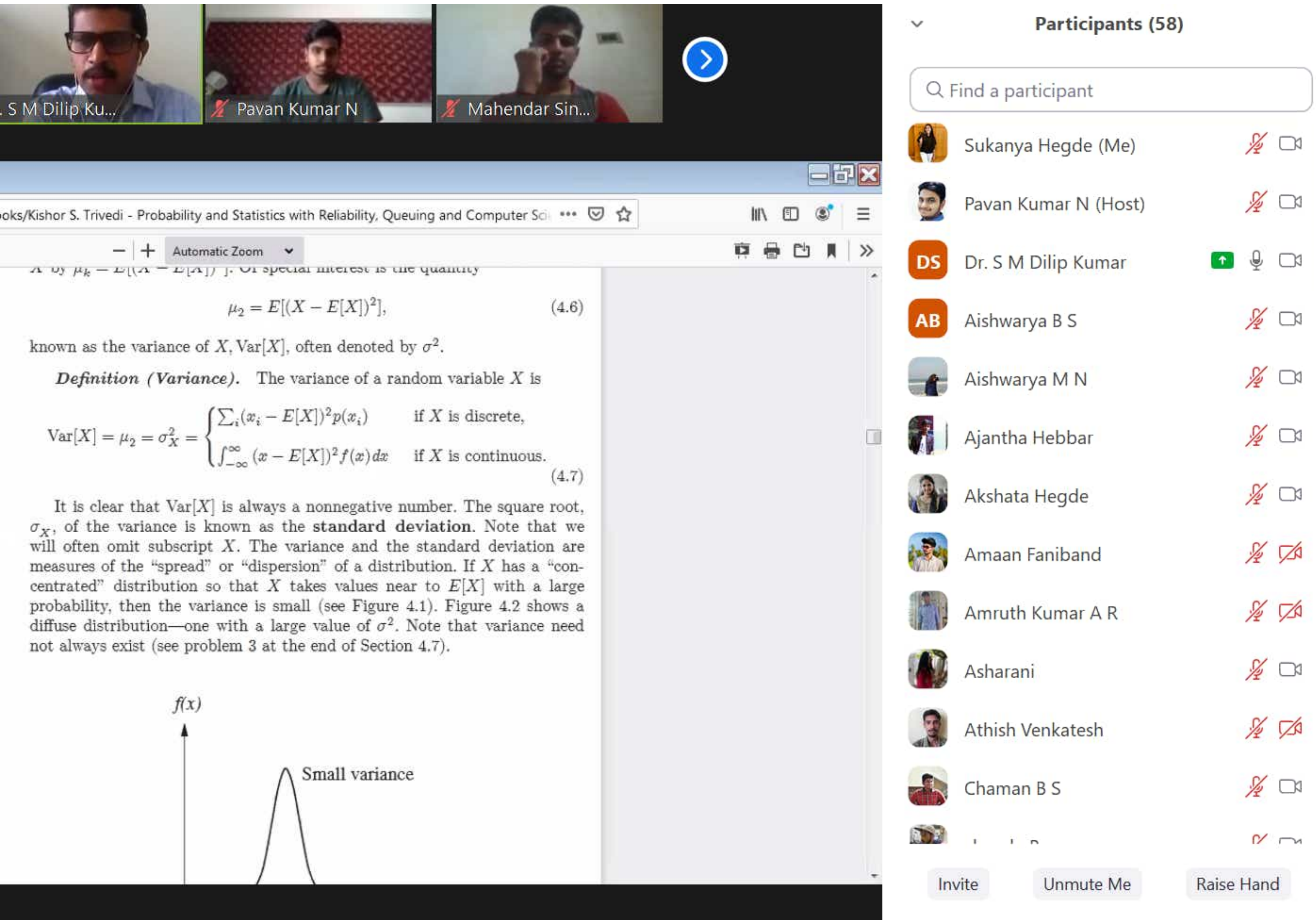
4. If X is uniformly distributed over $(0, 10)$, find the probability that
   (a) $X < 2$
   (b) $X > 2$

Using the formula for variance, we may write:

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{3(b-a)} \left[x^3\right]_a^b - \frac{(b+a)^2}{2} = \frac{b^3 - a^3}{3(b-a)} - \Big(\frac{b}{} \tag{7.31}$$

7.8.5 Exercises on uniform distribution

1. Arrivals of customers at a certain checkout counter follow a Poisson distribution. It is known that, during a given 30-minute period, one customer arrived at the counter. Find the probability that the customer arrived during the last 5 minutes of the 30-minute period.
   Solution: The actual time of arrival follows a uniform distribution over the interval of $(0,30)$. If X denotes the arrival time, then
   $P(25 \leq X \leq 30) = \int_{25}^{30} = \frac{1}{30} dx = \frac{30-25}{30} = \frac{1}{6}$.
2. Electric trains on a certain line run every half an hour between mid-night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least twenty minutes?
   Solution: Let the random variable X denote the waiting time in minutes for the next train. Given that a man arrives at the station at random, X is uniform distributed on $(0,30)$ with density
   $f(x) = \begin{cases} \frac{1}{30} & 0 < x < 30 \\ 0 & x \geq 30 \end{cases}$
   Thus the probability that he has to wait for at least 20 minutes is
   $P(X \geq 20) = \int_{20}^{30} f(x)dx = \int_{20}^{30} \frac{1}{30} dx = \frac{1}{30} \int_{20}^{30} dx = \frac{1}{30} [x]_{20}^{30} = \frac{1}{30} [30-20] = \frac{10}{30} = \frac{1}{3}$.
3. The time (in minutes) passenger must wait for a computer plane in a main railway station is uniformly distributed on the interval [0,60]. What is the probability that a passenger waits



$$\mu_2 = E[(X - E[X])^2], \tag{4.6}$$

known as the variance of X, $\mathrm{Var}[X]$, often denoted by σ^2.

Definition (Variance). The variance of a random variable X is

$$\mathrm{Var}[X] = \mu_2 = \sigma_X^2 = \begin{cases} \sum_i (x_i - E[X])^2 p(x_i) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} (x - E[X])^2 f(x)\,dx & \text{if } X \text{ is continuous.} \end{cases} \tag{4.7}$$

It is clear that $\mathrm{Var}[X]$ is always a nonnegative number. The square root, σ_X, of the variance is known as the **standard deviation**. Note that we will often omit subscript X. The variance and the standard deviation are measures of the "spread" or "dispersion" of a distribution. If X has a "concentrated" distribution so that X takes values near to $E[X]$ with a large probability, then the variance is small (see Figure 4.1). Figure 4.2 shows a diffuse distribution—one with a large value of σ^2. Note that variance need not always exist (see problem 3 at the end of Section 4.7).

3.3 THE RELIABILITY AND FAILURE RATE

Let the random variable X be the lifetime or the time to failure of a component. The probability that the component survives until some time t is called the **reliability** $R(t)$ of the component. Thus, $R(t) = P(X > t) = 1 - F(t)$, where F is the distribution function of the component lifetime X. The component is normally (but not always) assumed to be working properly at time $t = 0$ [i.e., $R(0) = 1$], and no component can work forever without failure [i.e., $\lim_{t \to +\infty} R(t) = 0$]. Also, $R(t)$ is a monotone decreasing function of t. For t less than zero, reliability has no meaning, but we let $R(t) = 1$ for $t < 0$. $F(t)$ will often be called the **unreliability**.

Consider a fixed number of identical components, N_0, under test. After time t, $N_f(t)$ components have failed and $N_s(t)$ components have survived with $N_f(t) + N_s(t) = N_0$. The estimated probability of survival may be written (using the frequency interpretation of probability) as

$$\hat{P}(\text{survival}) = \frac{N_s(t)}{N_0}.$$

In the limit as $N_0 \to \infty$, we expect $\hat{P}(\text{survival})$ to approach $R(t)$. As the test progresses, $N_s(t)$ gets smaller and $R(t)$ decreases:

$$\begin{aligned} R(t) &\simeq \frac{N_s(t)}{N_0} \\ &= \frac{N_0 - N_f(t)}{N_0} \end{aligned}$$

written (using the frequency interpretation of probability) as

$$\hat{P}(\text{survival}) = \frac{N_s(t)}{N_0}.$$

In the limit as $N_0 \to \infty$, we expect $\hat{P}(\text{survival})$ to approach $R(t)$. As the test progresses, $N_s(t)$ gets smaller and $R(t)$ decreases:

$$\begin{aligned} R(t) &\simeq \frac{N_s(t)}{N_0} \\ &= \frac{N_0 - N_f(t)}{N_0} \\ &= 1 - \frac{N_f(t)}{N_0}. \end{aligned}$$

The total number of components N_0 is constant, while the number of failed components N_f increases with time. Taking derivatives on both sides of the preceding equation, we get

$$R'(t) \simeq -\frac{1}{N_0} N_f'(t). \tag{3.6}$$

In this equation, $N_f'(t)$ is the rate at which components fail. Therefore, as $N_0 \to \infty$, the right-hand side may be interpreted as the negative of the failure density function, $f_X(t)$:

$$R'(t) = -f_X(t). \tag{3.7}$$

In the limit as $N_0 \to \infty$, we expect $P(\text{survival})$ to approach $R(t)$. As the test progresses, $N_s(t)$ gets smaller and $R(t)$ decreases:

$$\begin{aligned} R(t) &\simeq \frac{N_s(t)}{N_0} \\ &= \frac{N_0 - N_f(t)}{N_0} \\ &= 1 - \frac{N_f(t)}{N_0}. \end{aligned}$$

The total number of components N_0 is constant, while the number of failed components N_f increases with time. Taking derivatives on both sides of the preceding equation, we get

$$R'(t) \simeq -\frac{1}{N_0} N_f'(t). \tag{3.6}$$

In this equation, $N_f'(t)$ is the rate at which components fail. Therefore, as $N_0 \to \infty$, the right-hand side may be interpreted as the negative of the failure density function, $f_X(t)$:

$$R'(t) = -f_X(t). \tag{3.7}$$

Note that $f(t)\Delta t$ is the (unconditional) probability that a component will fail in the interval $(t, t + \Delta t]$. However, if we have observed the component functioning up to some time t, we expect the (conditional) probability of its failure to be different from $f(t)\Delta t$. This leads us to the notion of instantaneous failure rate as follows.

progresses, $N_s(t)$ gets smaller and $R(t)$ decreases.

$$\begin{aligned} R(t) &\simeq \frac{N_s(t)}{N_0} \\ &= \frac{N_0 - N_f(t)}{N_0} \\ &= 1 - \frac{N_f(t)}{N_0}. \end{aligned}$$

The total number of components N_0 is constant, while the number of failed components N_f increases with time. Taking derivatives on both sides of the preceding equation, we get

$$R'(t) \simeq -\frac{1}{N_0} N_f'(t). \tag{3.6}$$

In this equation, $N_f'(t)$ is the rate at which components fail. Therefore, as $N_0 \to \infty$, the right-hand side may be interpreted as the negative of the failure density function, $f_X(t)$:

$$R'(t) = -f_X(t). \tag{3.7}$$

Note that $f(t)\Delta t$ is the (unconditional) probability that a component will fail in the interval $(t, t + \Delta t]$. However, if we have observed the component functioning up to some time t, we expect the (conditional) probability of its failure to be different from $f(t)\Delta t$. This leads us to the notion of instantaneous failure rate as follows.

Remaining Meeting Time: 09:39

S. Trivedi - Probability and Statistics with Reliability, Queuing and Computer Science Applications-Wiley (2016).pdf - Adobe Reader

View Window Help

145 / 859 200%

Notice that the conditional probability that the component does not survive for an (additional) interval of duration x given that it has survived until time t can be written as

$$G_Y(x|t) = \frac{P(t < X \leq t+x)}{P(X > t)} = \frac{F(t+x) - F(t)}{R(t)}. \tag{3.8}$$

Definition (Instantaneous Failure Rate). The instantaneous failure rate $h(t)$ at time t is defined to be

$$h(t) = \lim_{x \to 0} \frac{1}{x} \frac{F(t+x) - F(t)}{R(t)} = \lim_{x \to 0} \frac{R(t) - R(t+x)}{xR(t)},$$

so that

$$h(t) = \frac{f(t)}{R(t)}. \tag{3.9}$$

Thus, $h(t)\Delta t$ represents the conditional probability that a component having survived to age t will fail in the interval $(t, t+\Delta t]$. Alternate terms for $h(t)$ are *hazard rate*, *force of mortality*, *intensity rate*, *conditional failure rate*, or simply *failure rate*. Failure rates in practice are so small that expressing them as failures per hour is not appropriate. Often, $h(t)$ is expressed in failures per 10,000 hours. Another commonly used unit is FIT (failures in time), which expresses failures per 10^9 or a billion hours.

It should be noted that the exponential distribution is characterized by a constant failure rate, since

mute Start Video 66 Participants Chat Share Screen Record

Remaining Meeting Time: 07:55

r S. Trivedi - Probability and Statistics with Reliability, Queuing and Computer Science Applications-Wiley (2016).pdf - Adobe Reader

lit View Window Help

Open 145 / 859 200%

Notice that the conditional probability that the component does not survive for an (additional) interval of duration x given that it has survived until time t can be written as

$$G_Y(x|t) = \frac{P(t < X \le t + x)}{P(X > t)} = \frac{F(t+x) - F(t)}{R(t)}. \tag{3.8}$$

Definition (Instantaneous Failure Rate). The instantaneous failure rate $h(t)$ at time t is defined to be

$$h(t) = \lim_{x \to 0} \frac{1}{x} \frac{F(t+x) - F(t)}{R(t)} = \lim_{x \to 0} \frac{R(t) - R(t+x)}{xR(t)},$$

so that

$$h(t) = \frac{f(t)}{R(t)}. \tag{3.9}$$

Thus, $h(t)\Delta t$ represents the conditional probability that a component having survived to age t will fail in the interval $(t, t + \Delta t]$. Alternate terms for $h(t)$ are *hazard rate*, *force of mortality*, *intensity rate*, *conditional failure rate*, or simply *failure rate*. Failure rates in practice are so small that expressing them as failures per hour is not appropriate. Often, $h(t)$ is expressed in failures per 10,000 hours. Another commonly used unit is FIT (failures in time), which expresses failures per 10^9 or a billion hours.

It should be noted that the exponential distribution is characterized by a constant failure rate, since

nmute Start Video 65 Participants Chat Share Screen Recor

Remaining Meeting Time: 07:15

S. Trivedi - Probability and Statistics with Reliability, Queuing and Computer Science Applications-Wiley (2016).pdf - Adobe Reader

View Window Help

145 / 859 200%

so that

$$h(t) = \frac{f(t)}{R(t)}. \tag{3.9}$$

Thus, $h(t)\Delta t$ represents the conditional probability that a component having survived to age t will fail in the interval $(t, t + \Delta t]$. Alternate terms for $h(t)$ are *hazard rate, force of mortality, intensity rate, conditional failure rate*, or simply *failure rate*. Failure rates in practice are so small that expressing them as failures per hour is not appropriate. Often, $h(t)$ is expressed in failures per 10,000 hours. Another commonly used unit is FIT (failures in time), which expresses failures per 10^9 or a billion hours.

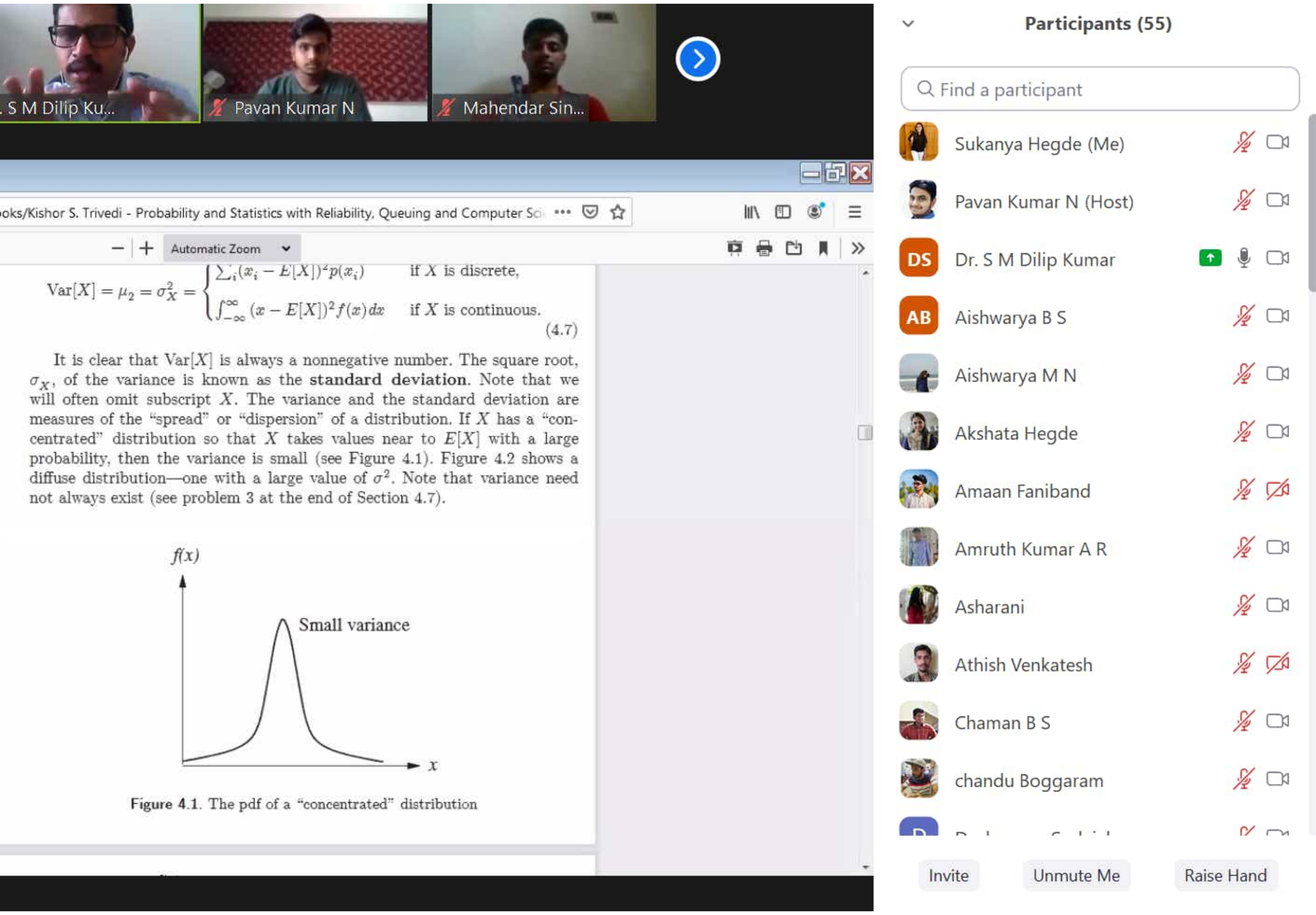
It should be noted that the exponential distribution is characterized by a constant failure rate, since

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

By integrating both sides of equation (3.9), we get

$$\begin{aligned}\int_0^t h(x)\,dx &= \int_0^t \frac{f(x)}{R(x)}\,dx\\ &= \int_0^t -\frac{R'(x)}{R(x)}\,dx \qquad \text{using equation (3.7)}\\ &= -\int_{R(0)}^{R(t)} \frac{dR}{R},\end{aligned}$$

mute Start Video 65 Participants Chat Share Screen Record



$$\mathrm{Var}[X] = \mu_2 = \sigma_X^2 = \begin{cases} \sum_i (x_i - E[X])^2 p(x_i) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} (x - E[X])^2 f(x)\,dx & \text{if } X \text{ is continuous.} \end{cases} \tag{4.7}$$

It is clear that Var[X] is always a nonnegative number. The square root, σ_X, of the variance is known as the **standard deviation**. Note that we will often omit subscript X. The variance and the standard deviation are measures of the "spread" or "dispersion" of a distribution. If X has a "concentrated" distribution so that X takes values near to $E[X]$ with a large probability, then the variance is small (see Figure 4.1). Figure 4.2 shows a diffuse distribution—one with a large value of σ^2. Note that variance need not always exist (see problem 3 at the end of Section 4.7).

Figure 4.1. The pdf of a "concentrated" distribution

as failures per hour is not appropriate. Often, $h(t)$ is expressed in failures per 10,000 hours. Another commonly used unit is FIT (failures in time), which expresses failures per 10^9 or a billion hours.

It should be noted that the exponential distribution is characterized by a constant failure rate, since

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

By integrating both sides of equation (3.9), we get

$$\begin{aligned} \int_0^t h(x)\,dx &= \int_0^t \frac{f(x)}{R(x)}\,dx \\ &= \int_0^t -\frac{R'(x)}{R(x)}\,dx \qquad \text{using equation (3.7)} \\ &= -\int_{R(0)}^{R(t)} \frac{dR}{R}, \end{aligned}$$

or

$$\int_0^t h(x)\,dx = -\ln R(t)$$

using the boundary condition, $R(0) = 1$. Therefore

$$R(t) = \exp\left[-\int_0^t h(x)\,dx\right]. \tag{3.10}$$

Remaining Meeting Time: 05:41

S. Trivedi - Probability and Statistics with Reliability, Queuing and Computer Science Applications-Wiley (2016).pdf - Adobe Reader

View Window Help

146 / 859 200%

This formula holds even when the distribution of the time to failure is not exponential.

The cumulative failure rate, $H(t) = \int_0^t h(x)\,dx$, is referred to as the **cumulative hazard**. Equation (3.10) gives a useful theoretical representation of reliability as a function of the failure rate. An alternate representation gives the reliability in terms of cumulative hazard:

$$R(t) = e^{-H(t)}. \tag{3.11}$$

Note that if the lifetime is exponentially distributed, then $H(t) = \lambda t$, and we obtain the exponential reliability function.

We should note the difference between $f(t)$ and $h(t)$. The quantity $f(t)\Delta t$ is the unconditional probability that the component will fail in the interval $(t, t+\Delta t]$, whereas $h(t)\Delta t$ is the conditional probability that the component will fail in the same time interval, given that it has survived until time t. Also, $h(t) = f(t)/R(t)$ is always greater than or equal to $f(t)$, because $R(t) \leq 1$. Function $f(t)$ represents probability density whereas $h(t)$ does not. By analogy, the probability that a newborn child will die at an age between 99 and 100 years [corresponding to $f(t)\Delta t$] is quite small because few of them will survive that long. But the probability of dying in that same period, provided that the child has survived until age 99 (corresponding to $h(t)\Delta t$) is much greater.

To further see the difference between the failure rate $h(t)$ and failure density $f(t)$, we need the notion of conditional probability density. Let $V_X(x|t)$ denote the conditional distribution of the lifetime X given that the component has survived past fixed time t. Then

mute Start Video 64 Participants Chat Share Screen Recor reliability as a function of the failure rate. An alternate representation gives the reliability in terms of cumulative hazard:

$$R(t) = e^{-H(t)}. \tag{3.11}$$

Note that if the lifetime is exponentially distributed, then $H(t) = \lambda t$, and we obtain the exponential reliability function.

We should note the difference between $f(t)$ and $h(t)$. The quantity $f(t)\Delta t$ is the unconditional probability that the component will fail in the interval $(t, t + \Delta t]$, whereas $h(t)\Delta t$ is the conditional probability that the component will fail in the same time interval, given that it has survived until time t. Also, $h(t) = f(t)/R(t)$ is always greater than or equal to $f(t)$, because $R(t) \leq 1$. Function $f(t)$ represents probability density whereas $h(t)$ does not. By analogy, the probability that a newborn child will die at an age between 99 and 100 years [corresponding to $f(t)\Delta t$] is quite small because few of them will survive that long. But the probability of dying in that same period, provided that the child has survived until age 99 (corresponding to $h(t)\Delta t$) is much greater.

To further see the difference between the failure rate $h(t)$ and failure density $f(t)$, we need the notion of conditional probability density. Let $V_X(x|t)$ denote the conditional distribution of the lifetime X given that the component has survived past fixed time t. Then

$$V_X(x|t) = \frac{\int_t^x f(y)\,dy}{P(X > t)}$$

$$\Bigl\{\ F(x) - F(t) \qquad x > t$$

reliability as a function of the failure rate. An alternate representation gives the reliability in terms of cumulative hazard:

$$R(t) = e^{-H(t)}. \tag{3.11}$$

Note that if the lifetime is exponentially distributed, then $H(t) = \lambda t$, and we obtain the exponential reliability function.

We should note the difference between $f(t)$ and $h(t)$. The quantity $f(t)\Delta t$ is the unconditional probability that the component will fail in the interval $(t, t + \Delta t]$, whereas $h(t)\Delta t$ is the conditional probability that the component will fail in the same time interval, given that it has survived until time t. Also, $h(t) = f(t)/R(t)$ is always greater than or equal to $f(t)$, because $R(t) \le 1$. Function $f(t)$ represents probability density whereas $h(t)$ does not. By analogy, the probability that a newborn child will die at an age between 99 and 100 years [corresponding to $f(t)\Delta t$] is quite small because few of them will survive that long. But the probability of dying in that same period, provided that the child has survived until age 99 (corresponding to $h(t)\Delta t$) is much greater.

To further see the difference between the failure rate $h(t)$ and failure density $f(t)$, we need the notion of conditional probability density. Let $V_X(x|t)$ denote the conditional distribution of the lifetime X given that the component has survived past fixed time t. Then

$$V_X(x|t) = \frac{\int_t^x f(y)\,dy}{P(X > t)}$$

$$\begin{cases} F(x) - F(t) & x > t \end{cases}$$

Then, using formula (3.12), we get

$$R(y+t) = R(t)\, R(y), \tag{3.13}$$

and rearranging, we get

$$\frac{R(y+t) - R(y)}{t} = \frac{[R(t) - 1]R(y)}{t}.$$

Taking the limit as t approaches zero and noting that $R(0) = 1$, we obtain

$$R'(y) = R'(0)R(y).$$

So $R(y) = e^{yR'(0)}$. Letting $R'(0) = -\lambda$, we get

$$R(y) = e^{-\lambda y}, \qquad y > 0,$$

which implies that the lifetime $X \sim EXP(\lambda)$. In this case, the failure rate $h(t)$ is equal to λ, which is a constant, independent of component age t. Conversely, the exponential lifetime distribution is the only distribution with a constant failure rate [BARL 1981]. If a component has exponential lifetime distribution, then it follows that

1. Since a used component is (stochastically) as good as new, a policy of a scheduled replacement of used components (known to be still functioning) does not accrue any benefit.
2. In estimating mean life, reliability, and other such quantities, data may be collected consisting only of the number of hours of observed life and of the number of observed failures; the ages of components under observation are of no concern

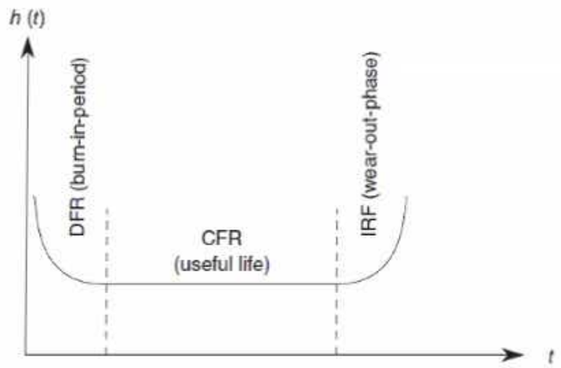


Figure 3.6. Failure rate as a function of time

is an increasing function of t for $t \geq 0$. The corresponding distribution function $F(t)$ is known as an **increasing failure rate (IFR) distribution**.

Alternately, if aging is beneficial in the sense that the conditional survival probability increases with age, then the failure rate will be a decreasing function of age, and the corresponding distribution is known as a **decreasing failure rate (DFR) distribution**.

The behavior of the failure rate as a function of age is known as the *mortality curve*, *hazard function*, *life characteristic*, or *lambda characteristic*. The mortality curve is empirically observed to have the so-called bathtub shape shown in Figure 3.6. During the early life period (infant mortality phase, burnin period, debugging period, or breakin period), failures are of the

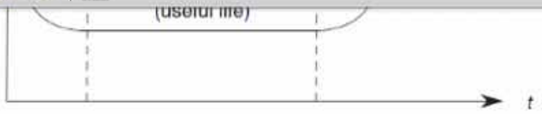


Figure 3.6. Failure rate as a function of time

is an increasing function of t for $t \geq 0$. The corresponding distribution function $F(t)$ is known as an **increasing failure rate (IFR) distribution**.

Alternately, if aging is beneficial in the sense that the conditional survival probability increases with age, then the failure rate will be a decreasing function of age, and the corresponding distribution is known as a **decreasing failure rate (DFR) distribution**.

The behavior of the failure rate as a function of age is known as the *mortality curve*, *hazard function*, *life characteristic*, or *lambda characteristic*. The mortality curve is empirically observed to have the so-called bathtub shape shown in Figure 3.6. During the early life period (infant mortality phase, burnin period, debugging period, or breakin period), failures are of the **endogenous** type and arise from inherent defects in the system attributed to faulty design, manufacturing, or assembly. During this period, the failure rate is expected to drop with age.

When the system has been debugged, it is prone to chance or random failure (also called **exogenous** failure). Such failures are usually associated with environmental conditions under which the component is operating. They are the results of severe, unpredictable stresses arising from sudden environmental shocks; the failure rate is determined by the severity of the environment is an increasing function of t for $t \geq 0$. The corresponding distribution function $F(t)$ is known as an **increasing failure rate (IFR) distribution**.

Alternately, if aging is beneficial in the sense that the conditional survival probability increases with age, then the failure rate will be a decreasing function of age, and the corresponding distribution is known as a **decreasing failure rate (DFR) distribution**.

The behavior of the failure rate as a function of age is known as the *mortality curve*, *hazard function*, *life characteristic*, or *lambda characteristic*. The mortality curve is empirically observed to have the so-called bathtub shape shown in Figure 3.6. During the early life period (infant mortality phase, burnin period, debugging period, or breakin period), failures are of the **endogenous** type and arise from inherent defects in the system attributed to faulty design, manufacturing, or assembly. During this period, the failure rate is expected to drop with age.

When the system has been debugged, it is prone to chance or random failure (also called **exogenous** failure). Such failures are usually associated with environmental conditions under which the component is operating. They are the results of severe, unpredictable stresses arising from sudden environmental shocks; the failure rate is determined by the severity of the environment. During this useful-life phase, failure rate is approximately constant and the exponential model is usually acceptable.

The rationale for the choice of exponential failure law is provided by assuming that the component is operating in an environment that subjects it to a stress varying in time. A failure occurs when the applied stress exceeds the maximum allowable stress S (see Figure 3.7). Such “peak” stresses may failure rate (DFR) distribution.

The behavior of the failure rate as a function of age is known as the *mortality curve, hazard function, life characteristic*, or *lambda characteristic*. The mortality curve is empirically observed to have the so-called bathtub shape shown in Figure 3.6. During the early life period (infant mortality phase, burnin period, debugging period, or breakin period), failures are of the **endogenous** type and arise from inherent defects in the system attributed to faulty design, manufacturing, or assembly. During this period, the failure rate is expected to drop with age.

When the system has been debugged, it is prone to chance or random failure (also called **exogenous** failure). Such failures are usually associated with environmental conditions under which the component is operating. They are the results of severe, unpredictable stresses arising from sudden environmental shocks; the failure rate is determined by the severity of the environment. During this useful-life phase, failure rate is approximately constant and the exponential model is usually acceptable.

The rationale for the choice of exponential failure law is provided by assuming that the component is operating in an environment that subjects it to a stress varying in time. A failure occurs when the applied stress exceeds the maximum allowable stress, $S_{\max}$ (see Figure 3.7). Such "peak" stresses may be assumed to follow a Poisson distribution with parameter λt, where λ is a constant rate of occurrence of peak loads. Denoting the number of peak stresses in the interval $(0, t]$ by N_t, we get

$$P(N_t = r) = \frac{e^{-\lambda t}(\lambda t)^r}{r!}, \qquad \lambda > 0, \quad r = 0, 1, 2, \ldots.$$

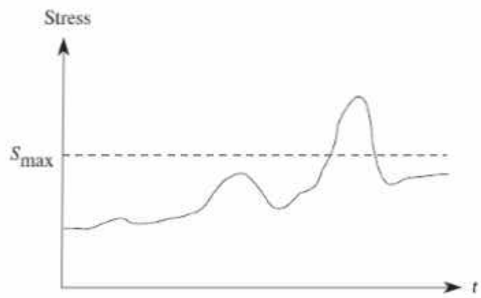


Figure 3.7. Stress as a function of time

Now the event $[X > t]$, where X is the component lifetime, corresponds to the event $[N_t = 0]$, and thus

$$\begin{aligned} R(t) &= P(X > t) \\ &= P(N_t = 0) \\ &= e^{-\lambda t}, \end{aligned}$$

the exponential reliability function.

When components begin to reach their "rated life," the system failure rate begins to increase and it is said to have entered the wearout phase. The wearout failure is the outcome of accumulated wear and tear, a result of a depletion process due to abrasion, fatigue, creep, and the like.

oks/Kishor S. Trivedi - Probability and Statistics with Reliability, Queuing and Computer Sci

Automatic Zoom

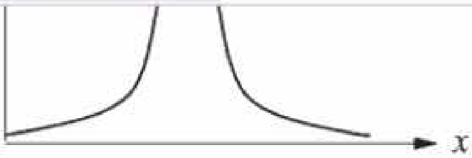


Figure 4.1. The pdf of a "concentrated" distribution

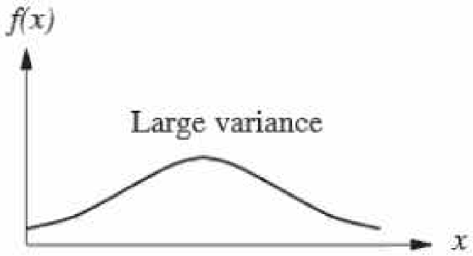


Figure 4.2. The pdf of a diffuse distribution

The third and the fourth central moments are called the *skewness* and *kurtosis*, respectively.

Example 4.6

Let X be an exponentially distributed random variable with parameter λ. Then, since $E[X] = 1/\lambda$, and $f(x) = \lambda e^{-\lambda x}$:

$$\sigma^2 = \int_0^\infty (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x}\,dx$$

$$= \int^\infty \lambda x^2 e^{-\lambda x}\,dx - 2\int^\infty x e^{-\lambda x}\,dx + \frac{1}{\lambda}\int^\infty e^{-\lambda x}\,dx$$

Figure 3.7. Stress as a function of time

Now the event $[X > t]$, where X is the component lifetime, corresponds to the event $[N_t = 0]$, and thus

$$
\begin{aligned}
R(t) &= P(X > t) \\
&= P(N_t = 0) \\
&= e^{-\lambda t},
\end{aligned}
$$

the exponential reliability function.

When components begin to reach their "rated life," the system failure rate begins to increase and it is said to have entered the wearout phase. The wearout failure is the outcome of accumulated wear and tear, a result of a depletion process due to abrasion, fatigue, creep, and the like.

Problems

1. The failure rate of a certain component is $h(t) = \lambda_0 t$, where $\lambda_0 > 0$ is a given constant. Determine the reliability, $R(t)$, of the component. Repeat for $h(t) = \lambda_0 t^{1/2}$.
2. The failure rate of a computer system for onboard control of a space vehicle is estimated to be the following function of time:

$$\begin{aligned} R(t) &= P(X > t) \\ &= P(N_t = 0) \\ &= e^{-\lambda t}, \end{aligned}$$

the exponential reliability function.

When components begin to reach their "rated life," the system failure rate begins to increase and it is said to have entered the wearout phase. The wearout failure is the outcome of accumulated wear and tear, a result of a depletion process due to abrasion, fatigue, creep, and the like.

Problems

1. The failure rate of a certain component is $h(t) = \lambda_0 t$, where $\lambda_0 > 0$ is a given constant. Determine the reliability, $R(t)$, of the component. Repeat for $h(t) = \lambda_0 t^{1/2}$.

2. The failure rate of a computer system for onboard control of a space vehicle is estimated to be the following function of time:

$$h(t) = \alpha\mu t^{\alpha-1} + \beta\gamma t^{\beta-1}.$$

Derive an expression for the reliability $R(t)$ of the system. Plot $h(t)$ and $R(t)$ as functions of time with parameter values $\alpha = \frac{1}{4}$, $\beta = \frac{1}{7}$, $\mu = 0.0004$, and $\gamma = 0.0007$.

definition (5.1) is adequate, as shown in the following definition.

Definition (Conditional pmf). Let X and Y be discrete random variables having a joint pmf $p(x, y)$. The conditional pmf of Y given X is

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y = y \mid X = x) \\ &= \frac{P(Y = y, X = x)}{P(X = x)} \qquad (5.2) \\ &= \frac{p(x,y)}{p_X(x)}, \end{aligned}$$

if $p_X(x) \neq 0$.

Note that the conditional pmf, as defined above, satisfies properties (pl)–(p3) of a pmf, discussed in Chapter 2. Rewriting the above definition another way, we have

$$p(x, y) = p_X(x)p_{Y|X}(y|x) = p_Y(y)p_{X|Y}(x|y). \qquad (5.3)$$

This is simply another form of the multiplication rule (of Chapter 1), and it gives us a way to compute the joint pmf regardless of whether X and Y are independent. *If X and Y are independent*, then from (5.3) and the definition of independence (in Chapter 2) we conclude that

$$p_{Y|X}(y|x) = p_Y(y). \qquad (5.4)$$

From (5.3) we also have the following marginal probability:

5.1 INTRODUCTION

We have seen that if two random variables are independent, then their joint distribution can be determined from their marginal distribution functions. In the case of dependent random variables, however, the joint distribution can not be determined in this simple fashion. This leads us to the notions of conditional pmf, conditional pdf, and conditional distribution.

Recalling the definition of conditional probability, $P(A|B)$, for two events A and B, we can define the **conditional probability** $P(A|X = x)$ of event A, given that the event $[X = x]$ has occurred, as

$$P(A|X = x) = \frac{P(A \text{ occurs and } X = x)}{P(X = x)} \tag{5.1}$$

whenever $P(X = x) \neq 0$. In Chapter 3 we noted that if X is a continuous random variable, then $P(X = x) = 0$ for all x. In this case, definition (5.1) is not satisfactory. On the other hand, if X is a discrete random variable, then definition (5.1) is adequate, as shown in the following definition.

Definition (Conditional pmf). Let X and Y be discrete random variables having a joint pmf $p(x, y)$. The conditional pmf of Y given X is

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y = y \mid X = x) \\ &= \frac{P(Y = y, X = x)}{P(X = x)} \end{aligned} \tag{5.2}$$

$$P(A|X=x) = \frac{P(A \text{ occurs and } X = x)}{P(X=x)} \tag{5.1}$$

whenever $P(X=x) \neq 0$. In Chapter 3 we noted that if X is a continuous random variable, then $P(X=x)=0$ for all x. In this case, definition (5.1) is not satisfactory. On the other hand, if X is a discrete random variable, then definition (5.1) is adequate, as shown in the following definition.

Definition (Conditional pmf). Let X and Y be discrete random variables having a joint pmf $p(x,y)$. The conditional pmf of Y given X is

$$\begin{aligned} p_{Y|X}(y|x) &= P(Y=y \mid X=x) \\ &= \frac{P(Y=y, X=x)}{P(X=x)} \qquad (5.2) \\ &= \frac{p(x,y)}{p_X(x)}, \end{aligned}$$

if $p_X(x) \neq 0$.

Note that the conditional pmf, as defined above, satisfies properties (pl)–(p3) of a pmf, discussed in Chapter 2. Rewriting the above definition another way, we have

$$p(x,y) = p_X(x) p_{Y|X}(y|x) = p_Y(y) p_{X|Y}(x|y). \tag{5.3}$$

This is simply another form of the multiplication rule (of Chapter 1), and it gives us a way to compute the joint pmf regardless of whether X and Y are independent. *If X and Y are independent*, then from (5.3) and the definition

$$= \frac{P(Y = y, X = x)}{P(X = x)} \tag{5.2}$$

$$= \frac{p(x, y)}{p_X(x)},$$

if $p_X(x) \neq 0$.

Note that the conditional pmf, as defined above, satisfies properties (p1)–(p3) of a pmf, discussed in Chapter 2. Rewriting the above definition another way, we have

$$p(x, y) = p_X(x) p_{Y|X}(y|x) = p_Y(y) p_{X|Y}(x|y). \tag{5.3}$$

This is simply another form of the multiplication rule (of Chapter 1), and it gives us a way to compute the joint pmf regardless of whether X and Y are independent. *If X and Y are independent*, then from (5.3) and the definition of independence (in Chapter 2) we conclude that

$$p_{Y|X}(y|x) = p_Y(y). \tag{5.4}$$

From (5.3) we also have the following marginal probability:

$$p_Y(y) = \sum_{\text{all } x} p(x, y) = \sum_{\text{all } x} p_{Y|X}(y|x) p_X(x). \tag{5.5}$$

This is another form of the theorem of total probability of Chapter 1.

We can also define the conditional distribution function $F_{Y|X}(y|x)$ of a

$$= \frac{p(x,y)}{p_X(x)},$$

if $p_X(x) \neq 0$.

Note that the conditional pmf, as defined above, satisfies properties (pl)–(p3) of a pmf, discussed in Chapter 2. Rewriting the above definition another way, we have

$$p(x,y) = p_X(x) p_{Y|X}(y|x) = p_Y(y) p_{X|Y}(x|y). \tag{5.3}$$

This is simply another form of the multiplication rule (of Chapter 1), and it gives us a way to compute the joint pmf regardless of whether X and Y are independent. *If X and Y are independent*, then from (5.3) and the definition of independence (in Chapter 2) we conclude that

$$p_{Y|X}(y|x) = p_Y(y). \tag{5.4}$$

From (5.3) we also have the following marginal probability:

$$p_Y(y) = \sum_{\text{all } x} p(x,y) = \sum_{\text{all } x} p_{Y|X}(y|x) p_X(x). \tag{5.5}$$

This is another form of the theorem of total probability of Chapter 1.

We can also define the conditional distribution function $F_{Y|X}(y|x)$ of a random variable Y, given a discrete random variable X by

$$F_{Y|X}(y|x) = P(Y \leq y \mid X = x) = \frac{P(Y \leq y \text{ and } X = x)}{P(X = x)} \tag{5.6}$$

for all values of y and for all values of x such that $P(X = x) > 0$. Definition (5.6) applies even for the case when Y is not discrete.

Note that the conditional distribution function can be obtained from the conditional pmf (assuming that both X and Y are discrete):

$$\sum_{t \leq y} p(x,t)$$

2. Recall the problem of the mischievous student trying to open a password-protected file, and determine the expected number of trials $E[N_n]$ and the variance $\text{Var}[N_n]$ for both techniques described in problem 3, Section 2.5.

3. The number of failures of a computer system in a week of operation has the following pmf:

No. of Failures	0	1	2	3	4	5	6
Probability	.18	.28	.25	.18	.06	.04	.01

(a) Find the expected number of failures in a week.
(b) Find the variance of the number of failures in a week.

4. In a Bell System study made in 1961 regarding the dialing of calls between White Plains, New York, and Sacramento, California, the pmf of the number of trunks, X, required for a connection was found to be

i	1	2	3	4	5
$p_X(i)$	.50	.30	.12	.07	.01

Determine the distribution function of X. Compute $E[X]$, $\text{Var}[X]$ and the mode of X. Let Y denote the number of telephone switching exchanges that this call has to pass through. Then $Y = X + 1$. Determine the pmf, the distribution function, the mean, and the variance of Y.

5. Let X, Y, and Z, respectively, denote EXP(1), two-stage hyperexponential with $\alpha_1 = .5 = \alpha_2$, $\lambda_1 = 2$, and $\lambda_2 = \frac{2}{3}$, and two-stage Erlang with parameter 2 random variables. Note that $E[X] = E[Y] = E[Z]$. Find the mode, the median, the variance, and the coefficient of variation of each random variable. Compare the densities of X, Y, and Z by plotting on the same graph. Similarly compare the three distribution functions.

6. Given a random variable X and two functions $h(x)$ and $g(x)$ satisfying the condition $h(x) \le g(x)$ for all x, show that

gives us a way to compute the joint pmf regardless of whether X and Y are independent. *If X and Y are independent*, then from (5.3) and the definition of independence (in Chapter 2) we conclude that

$$p_{Y|X}(y|x) = p_Y(y). \tag{5.4}$$

From (5.3) we also have the following marginal probability:

$$p_Y(y) = \sum_{\text{all } x} p(x,y) = \sum_{\text{all } x} p_{Y|X}(y|x)p_X(x). \tag{5.5}$$

This is another form of the theorem of total probability of Chapter 1.

We can also define the conditional distribution function $F_{Y|X}(y|x)$ of a random variable Y, given a discrete random variable X by

$$F_{Y|X}(y|x) = P(Y \le y \mid X = x) = \frac{P(Y \le y \text{ and } X = x)}{P(X = x)} \tag{5.6}$$

for all values of y and for all values of x such that $P(X = x) > 0$. Definition (5.6) applies even for the case when Y is not discrete.

Note that the conditional distribution function can be obtained from the conditional pmf (assuming that both X and Y are discrete):

$$F_{Y|X}(y|x) = \frac{\sum_{t \le y} p(x,t)}{p_X(x)} = \sum_{t \le y} p_{Y|X}(t|x). \tag{5.7}$$

Example 5.1

A server cluster has two servers labeled A and B. Incoming jobs are independently routed by the front end equipment (called server switch) to server A with probability p and to server B with probability $(1-p)$. The number of jobs, X, arriving per unit time is Poisson distributed with parameter λ. Determine the distribution function of the number of jobs, Y, received by server A, per unit time.

Let us determine the conditional probability of the event $[Y = k]$ given that event $[X = n]$ has occurred. Note that routing of the n jobs can be thought of as a

Definition (Conditional pdf). Let X and Y be continuous random variables with joint pdf $f(x, y)$. The conditional density $f_{Y|X}$ is defined by

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad \text{if } 0 < f_X(x) < \infty. \tag{5.8}$$

It can be easily verified that the function defined in (5.8) satisfies properties (f1) and (f2) of a pdf.

It follows from the definition of conditional density that

$$f(x, y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y). \tag{5.9}$$

This is the continuous analog of the multiplication rule (MR) of Chapter 1. If X and Y are independent, then

$$f(x, y) = f_X(x) f_Y(y),$$

which implies that

$$f_{Y|X}(y|x) = f_Y(y). \tag{5.10}$$

Conversely, if equation (5.10) holds, then it follows that X and Y are independent random variables. Thus (5.10) is a necessary and sufficient condition for two random variables X and Y, which have a joint density, to be independent.

From the expression of joint density (5.9), we can obtain an expression for the marginal density of Y in terms of conditional density by integration:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\, dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx. \end{aligned} \tag{5.11}$$

This is the continuous analog of the theorem of total probability.

Further, in the definition of conditional density, we can reverse the role of X and Y to define (whenever $f_Y(y) > 0$):

$$f(x, y) = f_X(x) f_Y(y),$$

which implies that

$$f_{Y|X}(y|x) = f_Y(y). \tag{5.10}$$

Conversely, if equation (5.10) holds, then it follows that X and Y are independent random variables. Thus (5.10) is a necessary and sufficient condition for two random variables X and Y, which have a joint density, to be independent.

From the expression of joint density (5.9), we can obtain an expression for the marginal density of Y in terms of conditional density by integration:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\, dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx. \end{aligned} \tag{5.11}$$

This is the continuous analog of the theorem of total probability.

Further, in the definition of conditional density, we can reverse the role of X and Y to define (whenever $f_Y(y) > 0$):

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

Using the expression (5.11) for $f_Y(y)$ and noting that $f(x, y) = f_X(x) f_{Y|X}(y|x)$, we obtain

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx}. \tag{5.12}$$

This is the continuous analog of Bayes' rule discussed in Chapter 1.

The conditional pdf can be used to obtain the conditional probability:

$$P(a \le Y \le b \,|\, X = x) = \int_a^b f_{Y|X}(y|x)\, dy, \quad a \le b. \tag{5.13}$$

$$f(x, y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y). \tag{5.9}$$

This is the continuous analog of the multiplication rule (MR) of Chapter 1. If X and Y are independent, then

$$f(x, y) = f_X(x) f_Y(y),$$

which implies that

$$f_{Y|X}(y|x) = f_Y(y). \tag{5.10}$$

Conversely, if equation (5.10) holds, then it follows that X and Y are independent random variables. Thus (5.10) is a necessary and sufficient condition for two random variables X and Y, which have a joint density, to be independent.

From the expression of joint density (5.9), we can obtain an expression for the marginal density of Y in terms of conditional density by integration:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\,dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\,dx. \end{aligned} \tag{5.11}$$

This is the continuous analog of the theorem of total probability.

Further, in the definition of conditional density, we can reverse the role of X and Y to define (whenever $f_Y(y) > 0$):

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

Using the expression (5.11) for $f_Y(y)$ and noting that $f(x, y) = f_X(x) f_{Y|X}(y|x)$, we obtain

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\,dx}. \tag{5.12}$$

This is the continuous analog of Bayes' rule discussed in Chapter 1.

The conditional pdf can be used to obtain the conditional probability:

$$f(x, y) = f_X(x) f_Y(y),$$

which implies that

$$f_{Y|X}(y|x) = f_Y(y). \tag{5.10}$$

Conversely, if equation (5.10) holds, then it follows that X and Y are independent random variables. Thus (5.10) is a necessary and sufficient condition for two random variables X and Y, which have a joint density, to be independent.

From the expression of joint density (5.9), we can obtain an expression for the marginal density of Y in terms of conditional density by integration:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\, dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx. \end{aligned} \tag{5.11}$$

This is the continuous analog of the theorem of total probability.

Further, in the definition of conditional density, we can reverse the role of X and Y to define (whenever $f_Y(y) > 0$):

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

Using the expression (5.11) for $f_Y(y)$ and noting that $f(x, y) = f_X(x) f_{Y|X}(y|x)$, we obtain

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx}. \tag{5.12}$$

This is the continuous analog of Bayes' rule discussed in Chapter 1.

The conditional pdf can be used to obtain the conditional probability:

$$P(a \le Y \le b \,|\, X = x) = \int_a^b f_{Y|X}(y|x)\, dy, \quad a \le b. \tag{5.13}$$

which implies that

$$f_{Y|X}(y|x) = f_Y(y). \tag{5.10}$$

Conversely, if equation (5.10) holds, then it follows that X and Y are independent random variables. Thus (5.10) is a necessary and sufficient condition for two random variables X and Y, which have a joint density, to be independent.

From the expression of joint density (5.9), we can obtain an expression for the marginal density of Y in terms of conditional density by integration:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\,dx. \end{aligned} \tag{5.11}$$

This is the continuous analog of the theorem of total probability.

Further, in the definition of conditional density, we can reverse the role of X and Y to define (whenever $f_Y(y) > 0$):

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Using the expression (5.11) for $f_Y(y)$ and noting that $f(x,y) = f_X(x)$ $f_{Y|X}(y|x)$, we obtain

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\,dx}. \tag{5.12}$$

This is the continuous analog of Bayes' rule discussed in Chapter 1.

The conditional pdf can be used to obtain the conditional probability:

$$P(a \le Y \le b \,|\, X = x) = \int_a^b f_{Y|X}(y|x)\,dy, \quad a \le b. \tag{5.13}$$

Using the expression (5.11) for $f_Y(y)$ and noting that $f(x,y) = f_X(x)$ $f_{Y|X}(y|x)$, we obtain

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x)\, dx}. \tag{5.12}$$

This is the continuous analog of Bayes' rule discussed in Chapter 1.

The conditional pdf can be used to obtain the conditional probability:

$$P(a \leq Y \leq b \,|\, X = x) = \int_a^b f_{Y|X}(y|x)\, dy, \quad a \leq b. \tag{5.13}$$

In particular, the conditional distribution function $F_{Y|X}(y|x)$ is *defined*, analogous to (5.6), as

$$\begin{aligned} F_{Y|X}(y|x) = P(Y \leq y \,|\, X = x) &= \frac{\int_{-\infty}^{y} f(x,t)\, dt}{f_X(x)} \\ &= \int_{-\infty}^{y} f_{Y|X}(t|x)\, dt. \end{aligned} \tag{5.14}$$

As motivation for definition (5.14) we observe that

$$\begin{aligned} F_{Y|X}(y|x) &= \lim_{h \to 0} P(Y \leq y \,|\, x \leq X \leq x+h) \\ &= \lim_{h \to 0} \frac{P(x \leq X \leq x+h \text{ and } Y \leq y)}{P(x \leq X \leq x+h)} \\ &= \lim_{h \to 0} \frac{\int_x^{x+h} \int_{-\infty}^{y} f(s,t)\, dt ds}{\int_x^{x+h} f_X(s)\, ds}. \end{aligned}$$

For some x_1^*, x_2^* with $x \leq x_1^*, x_2^* \leq x+h$, we obtain

In particular, the conditional distribution function $F_{Y|X}(y|x)$ is *defined*, analogous to (5.6), as

$$F_{Y|X}(y|x) = P(Y \leq y \mid X = x) = \frac{\int_{-\infty}^{y} f(x,t)\,dt}{f_X(x)}$$
$$= \int_{-\infty}^{y} f_{Y|X}(t|x)\,dt. \tag{5.14}$$

As motivation for definition (5.14) we observe that

$$F_{Y|X}(y|x) = \lim_{h\to 0} P(Y \leq y \mid x \leq X \leq x+h)$$
$$= \lim_{h\to 0} \frac{P(x \leq X \leq x+h \text{ and } Y \leq y)}{P(x \leq X \leq x+h)}$$
$$= \lim_{h\to 0} \frac{\int_{x}^{x+h}\int_{-\infty}^{y} f(s,t)\,dt\,ds}{\int_{x}^{x+h} f_X(s)\,ds}.$$

For some x_1^*, x_2^* with $x \leq x_1^*, x_2^* \leq x+h$, we obtain

$$F_{Y|X}(y|x) = \lim_{h\to 0} \frac{h\int_{-\infty}^{y} f(x_1^*,t)\,dt}{h f_X(x_2^*)}$$

(by the mean value theorem of integrals)

$$= \lim_{h\to 0} \frac{\int_{-\infty}^{y} f(x_1^*,t)\,dt}{f_X(x_2^*)}$$

(since both x_1^* and x_2^* approach x as h approaches 0)

$$= \int^{y} \frac{f(x,t)}{f_X(x)}\,dt$$

	1	2	3	4	5	6
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$

	1	2	3	4	5	6	$P_X(x) = f(x)$
0	0 $P(0,1)$	0 $P(0,1)$	$\frac{1}{32}$ $P(0,1)$	$\frac{2}{32}$ $P(0,1)$	$\frac{2}{32}$ $P(0,1)$	$\frac{3}{32}$ $P(0,1)$	$P(X=0)$ $=\frac{8}{32}$
1	$\frac{1}{16}$ $P(1,1)$	$\frac{1}{16}$ $P(1,1)$	$\frac{1}{8}$ $P(1,1)$	$\frac{1}{8}$ $P(1,1)$	$\frac{1}{8}$ $P(1,1)$	$\frac{1}{8}$ $P(1,1)$	$P(X=1)$ $=\frac{20}{32}$
2	$\frac{1}{32}$ $P(2,1)$	$\frac{1}{32}$ $P(2,1)$	$\frac{1}{64}$ $P(2,1)$	$\frac{1}{64}$ $P(2,1)$	0 $P(2,1)$	$\frac{2}{64}$ $P(2,1)$	$P(X=2)$ $=\frac{4}{32}$
$P_Y(y)$ $= f(y)$	$P(Y=1)$ $=\frac{3}{32}$	$P(Y=2)$ $=\frac{3}{32}$	$P(Y=3)$ $=\frac{11}{64}$	$P(Y=4)$ $=\frac{13}{64}$	$P(Y=5)$ $=\frac{6}{32}$	$P(Y=6)$ $=\frac{16}{64}$	1

8.8 Exercises on two-dimensional random variables

1. The joint probability mass function (pmf) of X and Y is given the table below: Compute the following:
   (a) $P(X \leq 1$
   (b) $P(Y \leq 3)$
   (c) $P(X \leq 1, Y \leq 3)$
   (d) $P(X \leq 1 | Y \leq 3)$
   (e) $P(Y \leq 3 | X \leq 1)$
   (f) $P(X+Y \leq 4)$

1	$\overline{16}$ $P(1,1)$	$\overline{16}$ $P(1,1)$	$\overline{8}$ $P(1,1)$	$\overline{8}$ $P(1,1)$	$\overline{8}$ $P(1,1)$	$\overline{8}$ $P(1,1)$	$P(X=1)$ $=\frac{20}{32}$
2	$\frac{1}{32}$ $P(2,1)$	$\frac{1}{32}$ $P(2,1)$	$\frac{1}{64}$ $P(2,1)$	$\frac{1}{64}$ $P(2,1)$	0 $P(2,1)$	$\frac{2}{64}$ $P(2,1)$	$P(X=2)$ $=\frac{4}{32}$
$P_Y(y)$ $=f(y)$	$P(Y=1)$ $=\frac{3}{32}$	$P(Y=2)$ $=\frac{3}{32}$	$P(Y=3)$ $=\frac{11}{64}$	$P(Y=4)$ $=\frac{13}{64}$	$P(Y=5)$ $=\frac{6}{32}$	$P(Y=6)$ $=\frac{16}{64}$	1

8.8 Exercises on two-dimensional random variables

1. The joint probability mass function (pmf) of X and Y is given the table below: Compute the following:
   (a) $P(X \leq 1$
   (b) $P(Y \leq 3)$
   (c) $P(X \leq 1, Y \leq 3)$
   (d) $P(X \leq 1 | Y \leq 3)$
   (e) $P(Y \leq 3 | X \leq 1)$
   (f) $P(X+Y \leq 4)$

   Solution: The given table can be re-written as follows:

   (a)
   $$P(X \leq 1) = P(X=0)+P(X=1)$$
   $$= \frac{8}{32}+\frac{20}{32} = \frac{28}{32} = \frac{7}{8}$$

   (b)
   $$P(Y \leq 3) = P(Y=1)+P(Y=2)+P(Y=3)$$
   $$= \frac{3}{32}+\frac{3}{32}+\frac{11}{64} = \frac{23}{64}$$

   (c)
   $$P(X \leq 1, Y \leq 3) = P(0,1)+P(0,2)+P(0,3)+P(1,1)+P(1,2)+P(1,3)$$
   $$= 0+0+\frac{1}{32}+\frac{1}{16}+\frac{1}{16}+\frac{1}{8} = \frac{9}{32}$$

$$E[h(X)] \leq E[g(X)],$$

whenever both expectations exist.

4.3 EXPECTATION BASED ON MULTIPLE RANDOM VARIABLES

Let $X_1, X_2, \ldots, X_n$ be n random variables defined on the same probability space, and let $Y = \phi(X_1, X_2, \ldots, X_n)$. Then

$$E[Y] = E[\phi(X_1, X_2, \ldots, X_n)]$$
$$= \begin{cases} \displaystyle\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n}\phi(x_1, x_2, \ldots, x_n)p(x_1, x_2, \ldots, x_n) \\ \qquad\qquad\qquad\qquad \text{(discrete case)}, \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\phi(x_1, x_2, \ldots, x_n)f(x_1, x_2, \ldots, x_n)dx_1 dx_2\cdots dx_n \\ \qquad\qquad\qquad\qquad \text{(continuous case)}. \end{cases} \tag{4.10}$$

Example 4.7

Consider a moving head disk with the innermost cylinder of radius a and the outermost cylinder of radius b. We assume that the number of cylinders is very large and the cylinders are very close to each other, so that we may assume a continuum of cylinders. Let the random variables X and Y, respectively, denote the current and the desired position of the head. Further assume that X and Y are independent and uniformly distributed over the interval (a, b). Therefore

$$f_X(x) = f_Y(y) = \frac{1}{b-a}, \quad a < x, y < b,$$

Compute the following:

(a) $P(X \leq 1$

(b) $P(Y \leq 3)$

(c) $P(X \leq 1, Y \leq 3)$

(d) $P(X \leq 1 | Y \leq 3)$

(e) $P(Y \leq 3 | X \leq 1)$

(f) $P(X + Y \leq 4)$

Solution: The given table can be re-written as follows:

(a)

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \frac{8}{32} + \frac{20}{32} = \frac{28}{32} = \frac{7}{8} \end{aligned}$$

(b)

$$\begin{aligned} P(Y \leq 3) &= P(Y = 1) + P(Y = 2) + P(Y = 3) \\ &= \frac{3}{32} + \frac{3}{32} + \frac{11}{64} = \frac{23}{64} \end{aligned}$$

(c)

$$\begin{aligned} P(X \leq 1, Y \leq 3) &= P(0,1) + P(0,2) + P(0,3) + P(1,1) + P(1,2) + P(1,3) \\ &= 0 + 0 + \frac{1}{32} + \frac{1}{16} + \frac{1}{16} + \frac{1}{8} = \frac{9}{32} \end{aligned}$$

(d)

$$\begin{aligned} P(X \leq 1 | Y \leq 3) &= \frac{P(X \leq 1, Y \leq 3)}{P(Y \leq 3)} \\ &= \frac{\frac{9}{32}}{\frac{23}{64}} = \frac{9}{32} \times \frac{64}{23} = \frac{18}{23} \end{aligned}$$

$$= 0 + 0 + \frac{1}{32} + \frac{1}{16} + \frac{1}{16} + \frac{1}{8} = \frac{9}{32}$$

(d)

$$\begin{aligned} P(X \le 1 | Y \le 3) &= \frac{P(X \le 1, Y \le 3)}{P(Y \le 3)} \\ &= \frac{\frac{9}{32}}{\frac{23}{64}} = \frac{9}{32} \times \frac{64}{23} = \frac{18}{23} \end{aligned}$$

	1	2	3
1	$\frac{1}{12}$	$\frac{1}{6}$	0
2	0	$\frac{1}{9}$	$\frac{1}{5}$
3	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$

	1	2	3	$p(y)$
1	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
2	0	$\frac{1}{9}$	$\frac{1}{5}$	$\frac{14}{45}$
3	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$	$\frac{79}{180}$
$p(x)$	$\frac{5}{36}$	$\frac{19}{36}$	$\frac{1}{3}$	1

(e)

$$\begin{aligned} P(Y \le 3 | X \le 1) &= \frac{P(X \le 1, Y \le 3)}{P(X \le 1)} \\ &= \frac{\frac{9}{32}}{\frac{7}{8}} = \frac{9}{32} \times \frac{8}{7} = \frac{9}{28} \end{aligned}$$

(f)

$$\begin{aligned} P(X + Y \le 4) &= P(0,1) + P(0,2) + P(0,3) + P(0,4) \\ &\quad + P(1,1) + P(1,2) \\ &\quad + P(1,3) + P(2,1) + P(2,2) \\ &= \frac{13}{32} \end{aligned}$$

	1	2	3
1	$\frac{1}{12}$	$\frac{1}{6}$	0
2	0	$\frac{1}{9}$	$\frac{1}{5}$
3	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$

	1	2	3	$p(y)$
1	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
2	0	$\frac{1}{9}$	$\frac{1}{5}$	$\frac{14}{45}$
3	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$	$\frac{79}{180}$
$p(x)$	$\frac{5}{36}$	$\frac{19}{36}$	$\frac{1}{3}$	1

(e)

$$
\begin{aligned}
P(Y \leq 3|X \leq 1) &= \frac{P(X \leq 1, Y \leq 3)}{P(X \leq 1)} \\
&= \frac{\frac{9}{32}}{\frac{7}{8}} = \frac{9}{32} \times \frac{8}{7} = \frac{9}{28}
\end{aligned}
$$

(f)

$$
\begin{aligned}
P(X+Y \leq 4) &= P(0,1) + P(0,2) + P(0,3) + P(0,4) \\
&\quad + P(1,1) + P(1,2) \\
&\quad + P(1,3) + P(2,1) + P(2,2) \\
&= \frac{13}{32}
\end{aligned}
$$

2. The following table represents the joint probability distribution of the discrete random variable (X,Y)
   (a) Evaluate the marginal distribution of X and Y.
   (b) Find the conditional distribution of X given $Y = 2$.
   (c) Find the conditional distribution of Y given $X = 3$.

   Solution: To compute the marginal distribution of X and Y, we re-write the given distribution table as follows:
   (a) The marginal distribution of X is: The marginal distribution of Y is:
   (b) Conditional distribution of X given $Y = 2$ is given

	1	2	3
1	$\frac{1}{12}$	$\frac{1}{6}$	0
2	0	$\frac{1}{9}$	$\frac{1}{5}$
3	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$

	1	2	3	$p(y)$
1	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
2	0	$\frac{1}{9}$	$\frac{1}{5}$	$\frac{14}{45}$
3	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$	$\frac{79}{180}$
$p(x)$	$\frac{5}{36}$	$\frac{19}{36}$	$\frac{1}{3}$	1

(e)

$$
\begin{aligned}
P(Y \leq 3|X \leq 1) &= \frac{P(X \leq 1, Y \leq 3)}{P(X \leq 1)} \\
&= \frac{\frac{9}{32}}{\frac{7}{8}} = \frac{9}{32} \times \frac{8}{7} = \frac{9}{28}
\end{aligned}
$$

(f)

$$
\begin{aligned}
P(X+Y \leq 4) &= P(0,1)+P(0,2)+P(0,3)+P(0,4) \\
&\quad +P(1,1)+P(1,2) \\
&\quad +P(1,3)+P(2,1)+P(2,2) \\
&= \frac{13}{32}
\end{aligned}
$$

2. The following table represents the joint probability distribution of the discrete random variable (X,Y)
   (a) Evaluate the marginal distribution of X and Y.
   (b) Find the conditional distribution of X given $Y = 2$.
   (c) Find the conditional distribution of Y given $X = 3$.

   Solution: To compute the marginal distribution of X and Y, we re-write the given distribution table as follows:
   (a) The marginal distribution of X is: The marginal distribution of Y is:
   (b) Conditional distribution of X given $Y = 2$ is given

$$P(Y=2|X=3)=\frac{P(Y=2\cap X=3)}{P(X=3)}=\frac{\frac{1}{5}}{\frac{1}{3}}=\frac{3}{5}$$

$$P(Y=3|X=3)=\frac{P(Y=3\cap X=3)}{P(X=3)}=\frac{\frac{2}{15}}{\frac{1}{3}}=\frac{2}{5}$$

3. The joint pdf of two random variables X and Y is given by

$$f(x,y)=\begin{cases}\frac{1}{8}x(x-y) & 0<x<2;-x<y<x\\ 0 & otherwise\end{cases}$$

Find $f(y|x)$.

Solution: From the definition of the conditional probability density function of Y from equation 8.11, we have

$f_{XY}(y|x)=\frac{f_{XY}(x,y)}{f_X(x)}$ where $f_X(x)$ is the marginal density function of X.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \int_{-x}^{x}\frac{1}{8}x(x-y)dy \\ &= \frac{1}{8}\left(x^2y-x\frac{y^2}{2}\right)_{-x}^{2}=\frac{1}{8}\left(x^3-\frac{x^3}{2}+x^3-\frac{x^3}{2}\right)=\frac{x^3}{8}, 0<x<2 \end{aligned}$$

$$f_{XY}(y|x)=\frac{\frac{1}{8}x(x-y)}{\frac{x^3}{8}}, 0<x<2;\ -x<y<x$$

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

4. The joint pdf of the random variable (X,Y) is given by

$f(x,y)=Kxye^{-(x^2+y^2)}; x>0, y>0.$

Find the value of K and prove that X and Y are independent

$f_{XY}(y|x) = \frac{\cdots}{\frac{x^3}{8}}, 0 < x < 2; -x < y < x$

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

4. The joint pdf of the random variable (X,Y) is given by

$f(x,y) = Kxye^{-(x^2+y^2)}; x > 0, y > 0.$

Find the value of K and prove that X and Y are independent.

Solution: Here the range space is the entire first quadrant of the xy–plane By the property of the joint pdf, we have

$\int_{x>0}\int_{y>0} Kxye^{-(x^2+y^2)}dxdy = 1$

$K\int_0^\infty\int_0^\infty ye^{-y^2}\cdot xe^{-x^2}dxdy = 1$

Put $x^2 = t$; so $2xdx = dt$

Then $\frac{K}{2}\int_0^\infty ye^{-y^2}\cdot\int_0^\infty e^{-t}dtdy = 1$

$\frac{K}{2}\int_0^\infty ye^{-y^2}\left(\frac{e^{-t}}{-1}\right)_0^\infty dy = 1$

$\frac{K}{2}\int_0^\infty ye^{-y^2}(1)dy = 1$

Put $y^2 = v$; so $2ydy = dv$

Then $\frac{K}{2}\cdot\frac{1}{2}\int_0^\infty e^{-v}dv = 1$

$\frac{K}{2}\cdot\frac{1}{2}\left(\frac{e^{-v}}{-1}\right)_0^\infty = 1$

$\frac{K}{2}\cdot\frac{1}{2}(1) = 1$

$\frac{K}{4} = 1 \Rightarrow K = 4$

To prove that X and Y are independent, we should show that $f(x,y) = f_X(x)f_Y(y)$ from equation 8.15. So we need to compute the marginal density of X and Y, i.e. $f_Y(x)$ and $f_Y(y)$

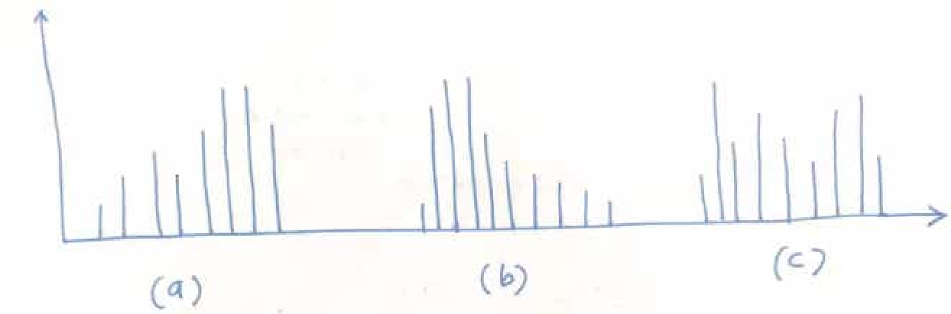
$$
\begin{aligned}
f_X(x) &= \int_0^\infty f(x,y)dy \\
&= 4x\int_0^\infty ye^{-(x^2+y^2)}dy \\
&= 4xe^{-x^2}\int_0^\infty ye^{-y^2}dy \\
&= 4xe^{-x^2}\int_0^\infty e^{-t}\frac{1}{2}dt \\
&= 2xe^{-x^2}\left[\frac{e^{-t}}{-1}\right]_0^\infty \\
&= 2xe^{-x^2}, x > 0
\end{aligned}
$$

Similarly

$$
\begin{aligned}
f_Y(y) &= \int_0^\infty f(x,y)dx \\
&= 4y\int_0^\infty ye^{-(x^2+y^2)}dx \\
&= 4ye^{-y^2}\int_0^\infty xe^{-x^2}dx \\
&= 4ye^{-y^2}\int_0^\infty e^{-t}\frac{1}{2}dt
\end{aligned}
$$

NORMAL DISTRIBUTION (Gaussian).

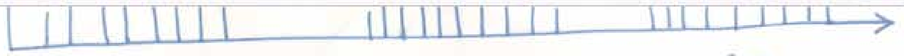
Data can be distributed (spread out) in different ways.



(a) spread out more on right

(b) ———— " ———— left

(c) jumbled up.

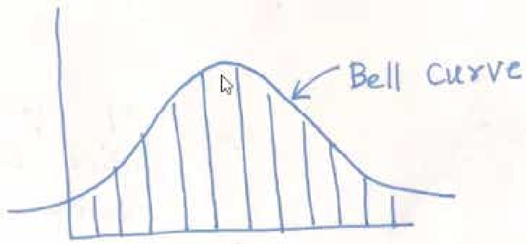


(a) (b) (c)

(a) spread out more on right

(b) ——— " ——— left

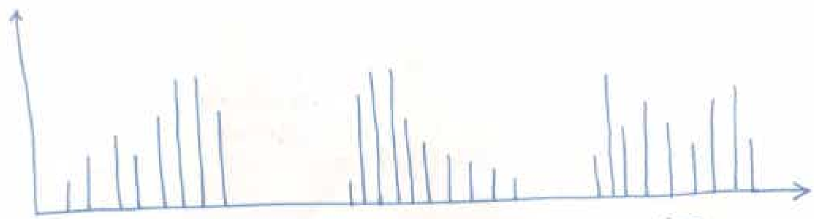
(c) jumbled up.



(d)

(d) Data tends to be around a central value with no bias left or right, and its gets close to a "normal distribution".

Data can be distributed (spread out) in different ways.

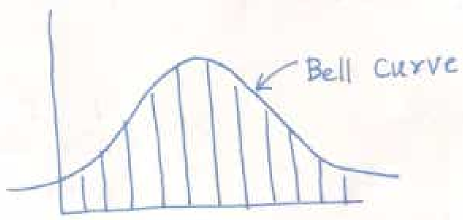


(a) (b) (c)

(a) spread out more on right

(b) ——— " ——— left

(c) jumbled up.



(d)

(d) Data tends to be around a central value with no bias left or right, and its gets close to a "normal distribution".

of the maximum. (In practice, the expected seek distance is even smaller because of correlations between successive requests [HUNT 1980, IBM 1997].)

Certain functions of random variables (e.g., sums), are of special interest and are of considerable use.

THEOREM 4.1 (The Linearity Property of Expectation). Let X and Y be two random variables. Then the expectation of their sum is the sum of their expectations; that is, if $Z = X + Y$, then $E[Z] = E[X+Y] = E[X] + E[Y]$.

Proof: We will prove the theorem assuming that X, Y, and hence Z are continuous random variables. Proof for the discrete case is very similar.

$$
\begin{aligned}
E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x,y)\,dy\,dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} x f_X(x)\,dx + \int_{-\infty}^{\infty} y f_Y(y)\,dy \\
&\quad \text{(by definition of the marginal densities)} \\
&= E[X] + E[Y].
\end{aligned}
$$

Note that Theorem 4.1 *does not* require that X and Y be independent. It

(d)

(d) Data tends to be around a central value with no bias left or right, and its gets close to a "normal distribution".

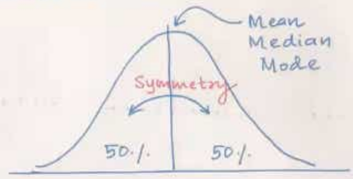
Ex: that follows normal distr. are

i) Heights of people

ii) errors in measurements

iii) Blood pressure

iv) Marks in a test.



50% of the values are < mean, and

50% —————"————— > mean.

Standard Deviations

(d)

(d) Data tends to be around a central value with no bias left or right, and its gets close to a "normal distribution".

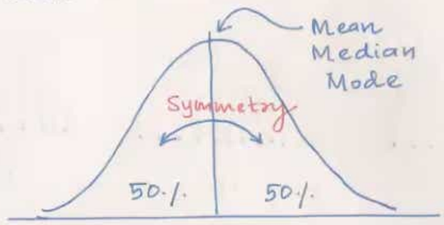
Ex: that follows normal distr. are

i) Heights of people

ii) errors in measurements

iii) Blood pressure

iv) Marks in a test.



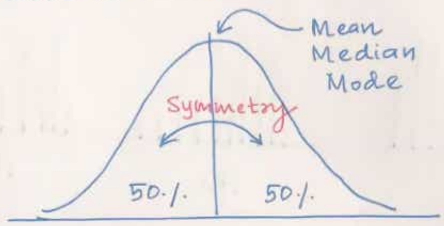
50% of the values are < mean, and
50% ——————— 〃 ——— > mean.

<u>Standard Deviations</u>

ii) errors in measur...

iii) Blood pressure

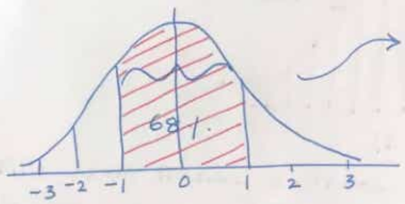
iv) Marks in a test.



50% of the values are < mean, and
50% ——"—— > mean.

Standard Deviations

S.D is a measure of how spread out numbers are

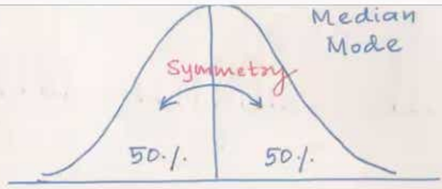


68% of values are within 1 SD of the mean.

or

95% – 2 SD

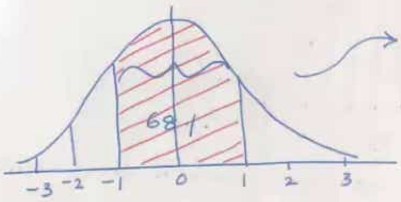
99.7% – 3 SD.



50% of the values are < mean, and
50% ——— " ——— > mean.

Standard Deviations

S.D is a measure of how spread out numbers are



68% of values are within 1SD of the mean.

or

95% – 2 SD

99.7% – 3 SD.

95% of students at school are b/w 1.1m

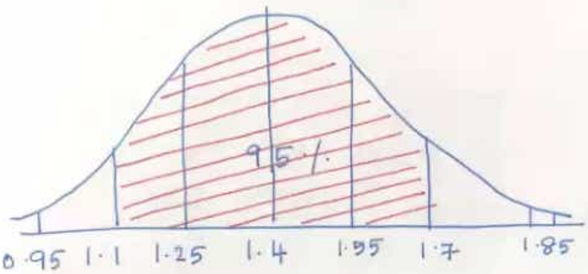
95% of students at school are b/w 1.1m and 1.7m tall. Assume the data is normally distributed, calculate mean and S.D.

Soln: Mean = (1.1 + 1.7)/2 = 1.4 m

95% of 2 SD either side of the mean (a total of 2SD) so

1 SD = (1.7 − 1.1)/4 = 0.15 m

As a result:



We say that any value is

* likely to be within 1 SD (68 out of 100 should be)
* very likely —"— 2SD (95/100)
* almost certainly —"— 3SD (997/1000).

The no. of S.D from the mean is also called the

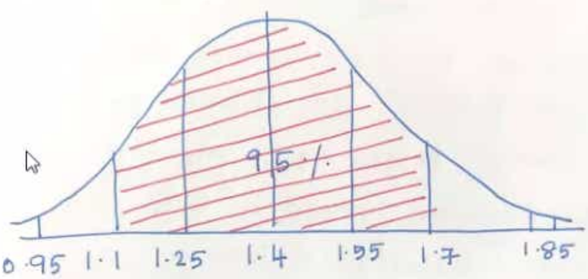
95%. of Students at school are b/w 1.1m and 1.7m tall. Assume the data is normally distributed, calculate mean and S.D.

Soln: Mean = (1.1+1.7)/2 = 1.4 m

95% of 2 SD either side of the mean (a total of 2SD) so

1 SD = (1.7−1.1)/4 = 0.15 m

As a result:



We say that any value is

* likely to be within 1 SD (68 out of 100 shouldbe)
* very likely —"— 2SD (95/100)
* almost certainly —"— 3SD (997/1000).

The no. of S.D from the mean is also called the

We say that any value is

* likely to be within 1 SD (68 out of 100 should be)
* very likely —"— 2 SD (95/100)
* almost certainly —"— 3 SD (997/1000).

The no. of S.D from the mean is also called the "Standard score", "sigma" or "z-score".

Ex: In that same school one of your friend is 1.85m tall.

1.85m is 3 SD from the mean 1.4, so: You friend height ~~is~~ has a z-score of 3.0.

or

$1.85 - 1.4 = 0.45\text{m}$ from the mean.

$1\,SD = 0.15$

$0.45 / 0.15 = 3.0\ S.D.$

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

The N.D is a prob. distr. fn. that describes how the values of a variable are distributed.

* likely to be within 1 SD (68 ...)
* very likely —"— 2 SD (95/100)
* almost certainly —"— 3 SD (997/1000).

The no. of S.D from the mean is also called the "Standard score", "sigma" or "z-score".

Ex: In that same school one of your friend is 1.85m tall.

1.85m is 3 SD from the mean 1.4, so: your friend height has a z-score of 3.0.

or

$1.85 - 1.4 = 0.45$ m from the mean.

$1 SD = 0.15$

$0.45 / 0.15 = 3.0$ S.D.

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

The N.D is a prob. distr. fn. that describes how the values of a variable are distributed. It is a symmetrical distr. where most of the observations cluster around the central peak and the probabilities for values further away friend height is has a Z-score of 3.0.

or

$1.85 - 1.4 = 0.45$ m from the mean.

1 SD = 0.15

$0.45 / 0.15 = 3.0$ S.D.

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

The N.D is a prob. distr. fn. that describes how the values of a variable are distributed.

It is a symmetrical distr. where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distr. are similarly unlikely.

Parameters of N.D.

The define its shape and probabilities entirely.

N.D has two parameters, the ~~shape~~ mean and S.D.

The shape changes based on parameters.

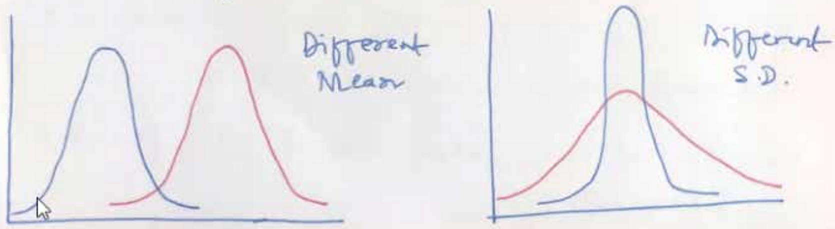
Different Mean

Different S.D.

and the probabilities for values further away from the mean taper off equally in both directions. Extreme values in both tails of the distr. are similarly unlikely.

Parameters of N.D.

The define its shape and probabilities entirely. N.D has two parameters, the ~~shape~~ mean and S.D. The shape changes based on parameters.



...ulation mean = μ

population S.D = σ

Sample mean = $\bar{X}$

Sample S.D = s