ECE 307 – Techniques for Engineering Decisions

12. Probability Distributions

George Gross

Department of Electrical and Computer Engineering

University of Illinois at Urbana-Champaign

SCOPE OF LECTURE

- We review basic probability distributions
- The entire lecture is simply a review of known probability material given the *course prerequisites*
- We extensively rely on examples to drive home the usefulness of the material
- We rely on the use of *probability distribution tables*, as found in the appendices of the *Clemen* book

OUTLINE OF DISTRIBUTION REVIEWED

- Discrete distributions
  - binomial
  - Poisson
- Continuous distributions
  - exponential
  - normal

THE BINOMIAL DISTRIBUTION

- Binomial distributions are used to describe events with only two possible outcomes
- Basic requirements are
  - *dichotomous outcomes*: uncertain events occur in a sequence with each event having one of two possible outcomes such as:

THE BINOMIAL DISTRIBUTION

- success/failure
- on/off
- correct/incorrect
- true/false

- *constant probability*: each event has the same probability of success
- *independence*: the outcome of each event is independent of the outcomes of any other event

BINOMIAL DISTRIBUTION EXAMPLE

- We consider a group of n identical machines with each machine having one of two states:

$$P\left\{ machine\ \ is\ \ on \right\} = p$$

$$P\left\{ machine\ is\ off \right\} = q = 1 - p$$

- For concreteness, let us set $n = 8$ and define for $i = 1, 2, \ldots, 8$, the $r.v.$s :

BINOMIAL DISTRIBUTION EXAMPLE

$$\underset{\sim}{X}_i = \begin{cases} 1 & \textit{machine } i \textit{ is on with prob.} \quad p \\ 0 & \textit{machine } i \textit{ is off with prob.} \quad q = 1 - p \end{cases}$$

- The probability that 3 or more machines are on is determined by the evaluation of the probability

$$P\left\{ \sum_{i=1}^{8} \underset{\sim}{X}_i \geq 3 \right\} = P\left\{ 3 \textit{ or more machines are on} \right\}$$

BINOMIAL DISTRIBUTION EXAMPLE

$$= P\left\{3 \ machines \ are \ on\right\} +$$

$$P\left\{4 \ machines \ are \ on\right\} +$$

$$\ldots \qquad +$$

$$P\left\{8 \ machines \ are \ on\right\}$$

$$P\left\{\sum_{i=1}^{8} \underset{\sim}{X}_i \geq 3\right\} = \sum_{r=3}^{8} \frac{8!}{\left(8-r\right)!r!} p^r \left(1-p\right)^{8-r}$$

THE BINOMIAL DISTRIBUTION

- In general, for a *r.v.* $\underset{\sim}{R}$ with *dichotomous outcomes* of success and failure, the probability of r successes in n trials is

$$P\left\{ \underset{\sim}{R} = r \ in \ n \ trials \ with \ probability \ of \ success \ p \right\}$$

$$= \frac{n!}{(n-r)!\,r!} p^{r} (1-p)^{n-r}$$

the binomial distribution

THE BINOMIAL DISTRIBUTION

- We can show that:

$$E\{\underset{\sim}{R}\} = n\,p$$

$$var\{\underset{\sim}{R}\} = n\,p\left(1 - p\right)$$

$$P\left\{\sum_{i=1}^{n} \underset{\sim}{X}_i \geq k\right\} = \sum_{r=k}^{n} \frac{n!}{\left(n-r\right)!\; r\,!}\, p^{r}\left(1 - \mathrm{p}\right)^{n-r}$$

EXAMPLE: SOFT PRETZELS

- A pretzel entrepreneur can sell each pretzel at *$* 0.50 with a market potential of 100,000 pretzels within a year; as there exists a competing product, he is not be the only seller
- Basic model is binomial:

new pretzel is a hit (*success*)	⇔	captures 30 % of market in one year
new pretzel is a flop (*failure*)	⇔	captures 10 % of market in one year

EXAMPLE: SOFT PRETZELS

- The probability of these two outcomes is equal
- Market tests are conducted with 20 pretzels taste tested against the competition; the result indicates that 5 out of 20 testers prefer the new pretzel
- We evaluate the conditional probability

$$P\left\{ new\ pretzel\ is\ a\ hit \,\middle|\, \mathbf{5}\ out\ of\ \mathbf{20}\ people\ prefer\ new\ pretzel \right\}$$

EXAMPLE: SOFT PRETZELS

- We define the success *r.v.*

$$\underset{\sim}{S} = \begin{cases} 1 & \textit{new pretzel is a hit (success)} \\ 0 & \textit{otherwise (failure)} \end{cases}$$

with

$$P\{\underset{\sim}{S} = 1\} = P\{\underset{\sim}{S} = 0\} = 0.5$$

and

$$\underset{\sim}{X}_i = \begin{cases} 1 & \textit{person } i \textit{ prefers new pretzel} \\ 0 & \textit{otherwise} \end{cases}$$

- We evaluate

$$P\{\textit{new pretzel is a hit} \mid 5 \textit{ out of } 20 \textit{ people prefer new pretzel}\}$$

EXAMPLE: SOFT PRETZELS

$$P\left\{ \underset{\sim}{S} = 1 \,\middle|\, \sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \right\} = \frac{P\left\{ \underset{\sim}{S} = 1, \sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \right\}}{P\left\{ \sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \right\}} =$$

$$\frac{P\left\{ \sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S} = 1 \right\} P\left\{ \underset{\sim}{S} = 1 \right\}}{P\left\{ \sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S} = 1 \right\} P\left\{ \underset{\sim}{S} = 1 \right\} + P\left\{ \sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S} = 0 \right\} P\left\{ \underset{\sim}{S} = 0 \right\}}$$

EXAMPLE: SOFT PRETZELS

$$\underbrace{P\left\{\sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S} = 1\right\}}_{}$$

0.178 from the binomial table

is the binomial probability that 5 out of 20 people prefer the new pretzel with $p = 0.3$

$$\underbrace{P\left\{\sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S} = 0\right\}}_{}$$

0.032 from the binomial table

is the binomial probability that 5 out of 20 people prefer the new pretzel with $p = 0.1$

EXAMPLE: SOFT PRETZELS

- Therefore,

$$\frac{P\left\{\sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S}=1\right\} P\left\{\underset{\sim}{S}=1\right\}}{P\left\{\sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S}=1\right\} P\left\{\underset{\sim}{S}=1\right\} + P\left\{\sum_{i=1}^{20} \underset{\sim}{X}_i = 5 \,\middle|\, \underset{\sim}{S}=0\right\} P\left\{\underset{\sim}{S}=0\right\}}$$

$$= \frac{(0.178)(0.5)}{(0.178)(0.5) + (0.032)(0.5)}$$

$$= 0.848$$

THE POISSON DISTRIBUTION

- The **binomial distribution** is appropriate for the representation of successes in repeated trials
- The **Poisson distribution** is appropriate for the representation of specific events over time, space, or some other problem–specific dimension, e.g., the number of customers who are served by a butcher in a meat market, or the number of chips judged unacceptable in a production run

REQUIREMENTS FOR A POISSON DISTRIBUTION

- **Events can happen at any of a large number of values within the range of measurement (time, space, etc.) and possibly along a continuum**

- **At a specific point z, P {*an event at* z} is very small and therefore events do not happen *too frequently***

REQUIREMENTS FOR A POISSON DISTRIBUTION

- **Each event is independent of any other event and**

$$P\left\{ event\ at\ any\ point \right\}$$

  is constant and *independent* of all other events

- **In fact, the average number of events over a unit of measure is constant**

THE POISSON DISTRIBUTED *r.v.*

- $\underset{\sim}{X}$ is the *r.v.* representing the number of events in a unit of measure

$$\left.\begin{array}{l} P\{\underset{\sim}{X} = k\} = \dfrac{e^{-m} m^{k}}{k!} \\ E\{\underset{\sim}{X}\} = m \quad var\{\underset{\sim}{X}\} = m \end{array}\right\} \text{ } m \text{ is the Poisson distribution parameter}$$

- Interpretation: the Poisson distribution parameter is the mean or the variance of the distribution

EXAMPLE: POISSON DISTRIBUTION

- Consider an assembly line for manufacturing a particular product
  - 1,024 units are produced
  - based on past experience, a flawed unit is pro–duced every 197 units and so, ***on average***, there are $\frac{1,024}{197} \text{»} 5.2$ flawed units in the 1,024 units of the product produced

EXAMPLE: POISSON DISTRIBUTION

- **Note that the Poisson conditions are satisfied**
  - **the sample has 1,024 units**
  - **there are only a few flawed units in the 1,024 sample, i.e., the event of the a flawed unit is infrequent**
  - **the probability of a flawed unit is rather small**
  - **each flawed unit is *independent* of every other flawed unit**

EXAMPLE: POISSON DISTRIBUTION

- Poisson distribution is appropriate representation

  with $m = 5.2$ and so,

$$P\{\underset{\sim}{X} = k\} = \frac{e^{-5.2}(5.2)^k}{k!}$$

- If we want to determine $P\{$4 or more flawed units$\}$,

  we compute

EXAMPLE: POISSON DISTRIBUTION

$$P\left\{\underset{\sim}{X} > 4\right\} = 1 - P\left\{\underset{\sim}{X} \leq 4\right\} = 1 - 0.406 = 0.594$$

lookup Poisson table for $k = 4$, $m = 5.2$

- **The Poisson table states that for $k = 12$, $m = 5.2$**

$$P\left\{\underset{\sim}{X} \leq 12\right\} = 0.997$$

and therefore

$$P\left\{\underset{\sim}{X} > 12\right\} = 1 - P\left\{\underset{\sim}{X} \leq 12\right\} = 0.003$$

EXAMPLE: SOFT PRETZELS

- The pretzel enterprise is going well: several retail outlets and a street vendor are selling the pretzels
- A vendor in a new location can sell, on average, 20 pretzels per hour; the vendor in an existing location sells 8 pretzels per hour

EXAMPLE: SOFT PRETZELS

- A decision is made to try to set up a second street vendor at a different, new location
- New location is classified along three distinct categories with the given probabilities

category	*characterization*	*probability*
"*good*"	20 *p/h are sold*	0.7
"*bad*"	10 *p/h are sold*	0.2
"*dismal*"	6 *p/h are sold*	0.1

EXAMPLE: SOFT PRETZELS

- After the first week, a long enough period to make a mark, a 30–minute test is run and 7 pretzels are sold during the 30–minute test period
- We analyze the situation and define the *r.v.*

$$\underset{\sim}{L} = \begin{cases} \text{"}good\text{"} & 10 \quad p.\ sold\ during\ test\ period \\ \text{"}bad\text{"} & 5 \quad p.\ sold\ during\ test\ period \\ \text{"}dismal\text{"} & 3 \quad p.\ sold\ during\ test\ period \end{cases}$$

and assume Poisson distribution applies

EXAMPLE: SOFT PRETZELS

- We determine the conditional probabilities of the new location conditioned on the 30-*minute* test outcomes and evaluate

$$P\left\{\underset{\sim}{L} = \text{"}good\text{"} \middle| \underset{\sim}{X} = 7\right\},\ P\left\{\underset{\sim}{L} = \text{"}bad\text{"} \middle| \underset{\sim}{X} = 7\right\}$$ and

$$P\left\{\underset{\sim}{L} = \text{"}dismal\text{"} \middle| \underset{\sim}{X} = 7\right\}$$

- We compute the values of the Poisson distributed

$$P\left\{\underset{\sim}{X} = 7 \middle| \underset{\sim}{L} = \text{"}good\text{"}\right\} = \frac{e^{-10}(10)^7}{7!} = 0.09$$

EXAMPLE: SOFT PRETZELS

$$P\{\underset{\sim}{X} = 7 \mid \underset{\sim}{L} = \text{" bad "}\} = \frac{e^{-5}(5)^7}{7!} = 0.104$$

$$P\{\underset{\sim}{X} = 7 \mid \underset{\sim}{L} = \text{" dismal "}\} = \frac{e^{-3}(3)^7}{7!} = 0.022$$

- Then $P\{\underset{\sim}{L} = \text{" good "} \mid \underset{\sim}{X} = 7\} =$

$$\frac{P\{\underset{\sim}{X} = 7 \mid \underset{\sim}{L} = \text{" good "}\} \cdot P\{\underset{\sim}{L} = \text{" good "}\}}{\left[\begin{array}{l} P\{\underset{\sim}{X} = 7 \mid \underset{\sim}{L} = \text{" good "}\} \cdot P\{\underset{\sim}{L} = \text{" good "}\} + P\{\underset{\sim}{X} = 7 \mid \underset{\sim}{L} = \text{" bad "}\} \\ \cdot P\{\underset{\sim}{L} = \text{" bad "}\} + P\{\underset{\sim}{X} = 7 \mid \underset{\sim}{L} = \text{" dismal "}\} \cdot P\{\underset{\sim}{L} = \text{" dismal "}\} \end{array}\right]}$$

EXAMPLE: SOFT PRETZELS

$$P\{\underset{\sim}{L} = "good" \mid \underset{\sim}{X} = 7\} = \frac{(0.09)(0.7)}{(0.09)(0.7)+(0.104)(0.2)+(0.022)(0.1)}$$

$$= 0.733$$

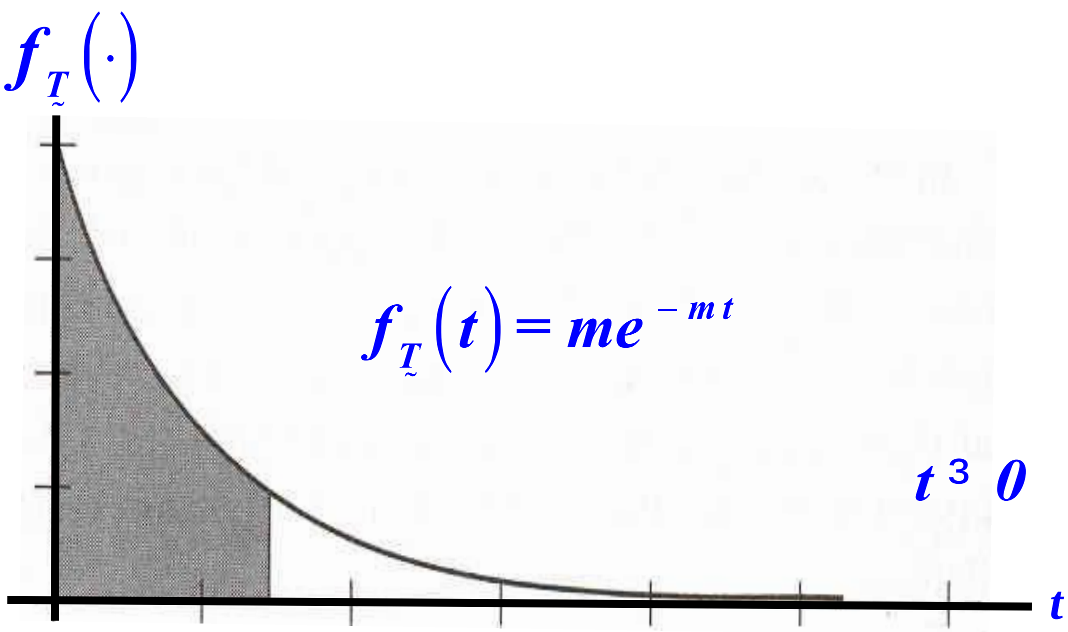
- Similarly,

$$P\left\{\underset{\sim}{L} = "bad" \quad \middle| \underset{\sim}{X} = 7\right\} = 0.242$$

$$P\left\{\underset{\sim}{L} = "dismal" \middle| \underset{\sim}{X} = 7\right\} = 1 - \left(0.733 + 0.242\right) = 0.025$$

EXPONENTIALLY DISTRIBUTED *r.v.*

- Unlike the discrete Poisson or the binomial distributed *r.v.*s, the exponentially distributed *r.v.* is continuous
- The density function has the form



EXPONENTIALLY DISTRIBUTED *r.v.*

- The exponentially distributed *r.v.* is related to the Poisson distribution in the following manner
- Consider the Poisson distributed *r.v.* $\underset{\sim}{X}$, where $\underset{\sim}{X}$ represents the number of events in a specified quantity of measure, *e.g.*, period of time
- We define $\underset{\sim}{T}$ to be the *r.v.* for the uncertain quantity we measure, *e.g.*, the time between 2 sequential events or the distance between 2 accidents

EXPONENTIALLY DISTRIBUTED *r.v.*

- Then, $\underset{\sim}{T}$ has the exponential distribution with

$$F_{\underset{\sim}{T}}(t) = P\{\underset{\sim}{T} \leq t\} = 1 - e^{-mt},$$

$$E\{\underset{\sim}{T}\} = \frac{1}{m} \quad \text{and} \quad var\{\underset{\sim}{T}\} = \frac{1}{m^2}$$

- The exponentially distributed *r.v.* is completely specified by the parameter m

EXAMPLE: SOFT PRETZELS

- We know that it takes $3.5\ minutes$ to bake a pretzel and we wish to determine the probability that the next customer will arrive after the pretzel baking is completed, i.e., $P\left\{ \underset{\sim}{T} > 3.5\ minutes \right\}$
- We also are given that the location types are classified as being

EXAMPLE: SOFT PRETZELS

"good" *location* $\Leftrightarrow$ $m = 20$ *pretzels / hour*

"bad" *location* $\Leftrightarrow$ $m = 10$ *pretzels / hour*

"dismal" *location* $\Leftrightarrow$ $m = 6$ *pretzels / hour*

- **We compute the probability by conditioning on the location type to obtain**

EXAMPLE: SOFT PRETZELS

$$P\left\{\underset{\sim}{T} > 3.5\,minutes\right\} = P\left\{\underset{\sim}{T} > 3.5\,minutes \,\middle|\, m = 20\right\} \cdot P\left\{m = 20\right\} +$$

$$P\left\{\underset{\sim}{T} > 3.5\,minutes \,\middle|\, m = 10\right\} \cdot P\left\{m = 10\right\} +$$

$$P\left\{\underset{\sim}{T} > 3.5\,minutes \,\middle|\, m = 6\right\} \cdot P\left\{m = 6\right\}$$

$\equiv 0.0583\,hour$

- We evaluate

$$P\left\{\underset{\sim}{T} > 3.5m\right\} =$$

EXAMPLE: SOFT PRETZELS

$$e^{-0.0583(20)}P\{m=20\}+e^{-0.0583(10)}P\{m=10\}+e^{-0.0583(6)}P\{m=6\}$$

ex post probabilities

$$P\{m=20\}=P\{\underset{\sim}{L}=\text{"good"}\,|\,\underset{\sim}{X}=7\}=0.733$$

$$P\{m=10\}=P\{\underset{\sim}{L}=\text{"bad"}\,|\,\underset{\sim}{X}=7\}=0.242$$

$$P\{m=6\}=P\{\underset{\sim}{L}=\text{"dismal"}\,|\,\underset{\sim}{X}=7\}=0.025$$

EXAMPLE: SOFT PRETZELS

and so

$$P\{\underset{\sim}{T} > 3.5 \; minutes\} = 0.3809$$

- Therefore,

$$P\{\underset{\sim}{T} \leq 3.5 \; minutes\} = 1 - 0.3809 = 0.6191$$

and the interpretation is that the majority of the customers arrives before the pretzels are baked

THE NORMAL DISTRIBUTION

- **The *normal* or *Gaussian* distribution is, by far, the most important probability distribution since the *Law of Large Numbers* implies that the distribution of many uncertain variables is governed by the *normal* distribution, or more commonly known as the *bell curve***
- **We consider a normally distributed *r.v.* $\underset{\sim}{Y}$**

$$\underset{\sim}{Y} \sim \mathcal{N}\left(\mu, \sigma \right)$$

THE NORMAL DISTRIBUTION

- The density function is

$$f_{\underset{\sim}{Y}}(y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\left(\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right)}$$

mean

variance

standard deviation

with $E\{\underset{\sim}{Y}\} = \mu$ *and* $var\{\underset{\sim}{Y}\} = \sigma^2$

THE STANDARD NORMAL DISTRIBUTION

- Consider the *r.v.* $\underset{\sim}{Z}$ which has the standard normal distribution

$$\underset{\sim}{Z} \sim \mathcal{N}(0,1)$$

1. The relationship between the *r.v.s* $\underset{\sim}{Y}$ and $\underset{\sim}{Z}$ is given by the affine expression:

$$\underset{\sim}{Z} = \frac{\underset{\sim}{Y} - \mu}{\sigma}$$

with

$$P\left\{\underset{\sim}{Y} \leq a\right\} = P\left\{\underset{\sim}{Z} \leq (a-\mu)/\sigma\right\}$$

THE STANDARD NORMAL DISTRIBUTION

- **Note that**

$$E\left\{\underset{\sim}{Z}\right\} = 0 \quad \textbf{and} \quad var\left\{\underset{\sim}{Z}\right\} = 1$$

- **In general, any value of the normal distribution is obtained from the *standard normal distribution* with the affine transformation**

$$\underset{\sim}{Z} = \frac{\underset{\sim}{Y} - \mu}{\sigma}$$

EXAMPLE: QUALITY CONTROL

- We consider a disk drive manufacturing process in which a particular machine produces a part used in the final assembly; the part must rigorously meet the width requirements to be within the interval $[3.995, 4.005]$ *mm* ; else, the company incurs $\$10.40$ in repair costs
- The machine is set to produce parts with the width of $4mm$, but in actual conditions, the width is a normally distributed *r.v.* $\underset{\sim}{W}$ with

EXAMPLE: QUALITY CONTROL

$$\underset{\sim}{W} \sim \mathcal{N}\left(4, \sigma\right)$$

and

$$\sigma = f\left(\textit{speed of machine}\right) = \begin{cases} 0.0019 & \textit{slow speed} \\ 0.0026 & \textit{high speed} \end{cases}$$

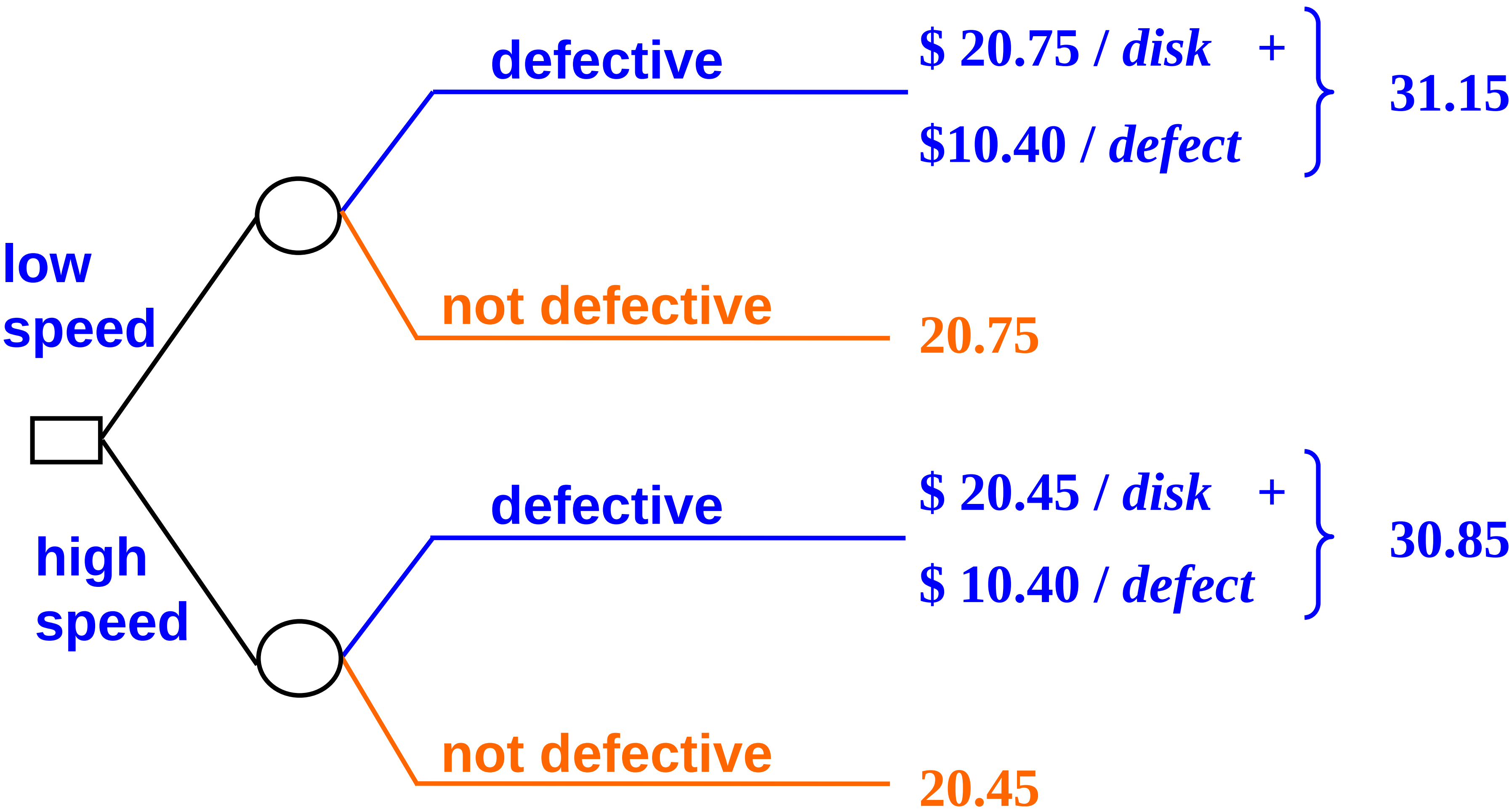
- The respective costs in *$* of the disk drive are

20.75	*slow speed*
20.45	*high speed*

EXAMPLE: QUALITY CONTROL

- We may interpret the cost data to imply that more disks can be produced at lower costs at the high speed
- The problem is to select the machine speed to obtain the more cost effective result
- A decision tree is useful in the analysis of the situation

EXAMPLE: QUALITY CONTROL



- We compute the probability of each outcome

LOW – SPEED PROBABILITY EVALUATION

$$P\left\{ defective\ disk\ is\ produced \right\} =$$

$$P\left\{ \underset{\sim}{W} < 3.995\ or\ \underset{\sim}{W} > 4.005 \right\} =$$

$$1 - P\left\{ 3.995 \leq \underset{\sim}{W} \leq 4.005 \right\} =$$

$$\underset{\sim}{Z} = \frac{\underset{\sim}{W} - 4}{0.0019}$$

$$1 - P\left\{ \frac{3.995 - 4}{0.0019} \leq \underset{\sim}{Z} \leq \frac{4.005 - 4}{0.0019} \right\} =$$

LOW – SPEED PROBABILITY EVALUATION

$$1 - P\left\{-2.63 \le \underset{\sim}{Z} \le 2.63\right\} \quad =$$

$$1 - \left[P\left\{\underset{\sim}{Z} \le 2.63\right\} - P\left\{\underset{\sim}{Z} \le -2.63\right\} \right] \quad = \quad 0.0086$$

0.9957 0.0043

0.9914

HIGH – SPEED PROBABILITY EVALUATION

$$P\left\{ \textit{defective disk is produced} \right\} =$$

$$P\left\{ \underset{\sim}{W} < 3.995 \textit{ or } \underset{\sim}{W} > 4.005 \right\} =$$

$$1 - P\left\{ 3.995 \leq \underset{\sim}{W} \leq 4.005 \right\} =$$

$$\underset{\sim}{Z} = \frac{\underset{\sim}{W} - 4}{0.0026}$$

$$1 - P\left\{ \frac{3.995 - 4}{0.0026} \leq \underset{\sim}{Z} \leq \frac{4.005 - 4}{0.0026} \right\} =$$

HIGH – SPEED PROBABILITY EVALUATION

$$1 - P\left\{-1.92 \leq \underset{\sim}{Z} \leq 1.92\right\} =$$

$$1 - \left[P\left\{\underset{\sim}{Z} \leq 1.92\right\} - P\left\{\underset{\sim}{Z} \leq -1.92\right\} \right] = 0.0548$$

0.9726

0.0274

0.9452

MEAN VALUE EVALUATION

- We next evaluate the mean cost per disk

$$E\left\{cost\,/\,disk \,\middle|\, low\ speed\right\} = \left(0.9914\right)\left(20.75\right) + \left(0.0086\right)\left(31.15\right)$$

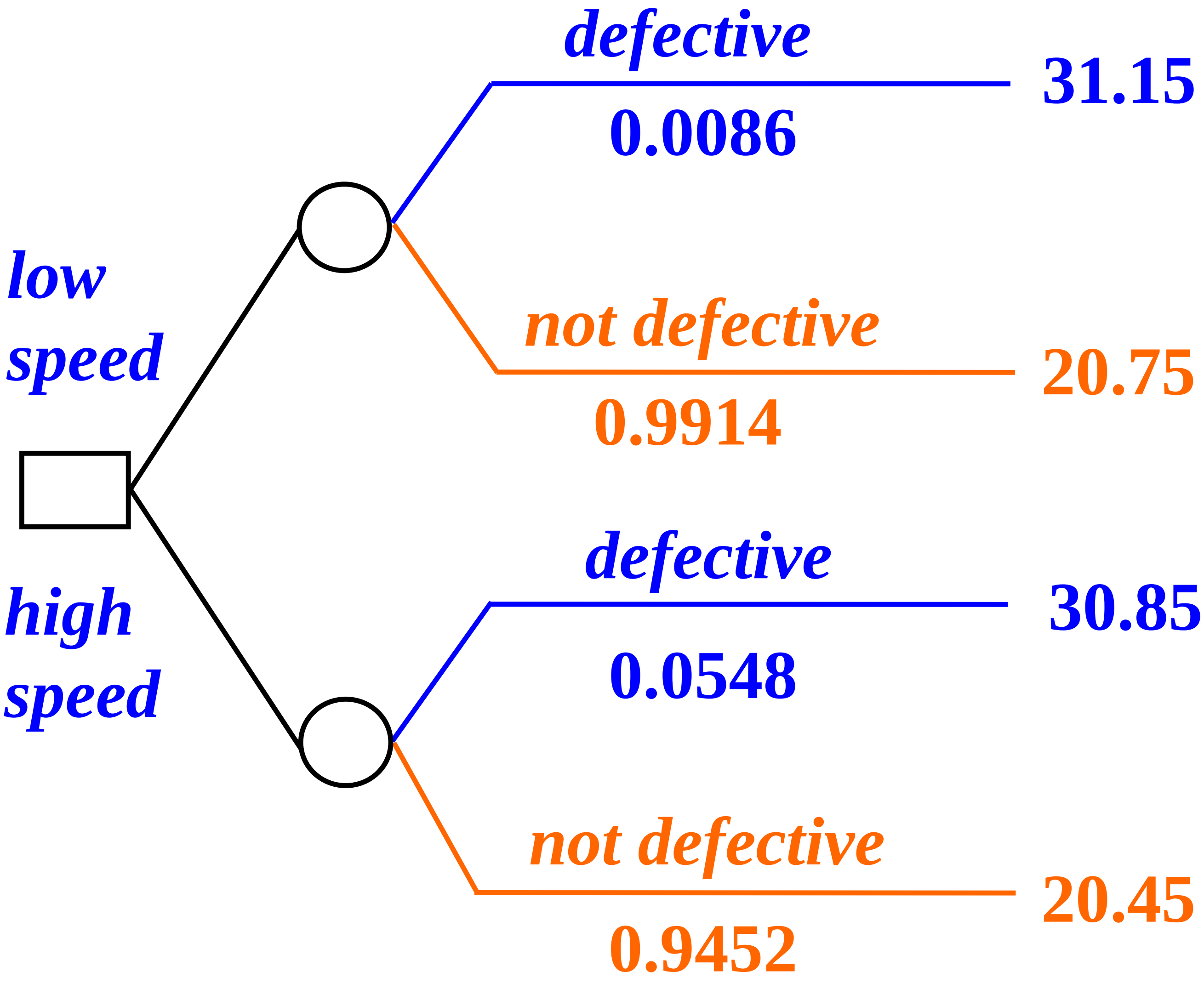
$$= 20.84$$

$$E\left\{cost\,/\,disk \,\middle|\, high\ speed\right\} = \left(0.9452\right)\left(20.45\right) + \left(0.0548\right)\left(30.85\right)$$

$$= 21.02$$

- We summarize the information in the decision tree

EXAMPLE: QUALITY CONTROL



$M_{low} = 20.84$

$M_{high} = 21.02$

CASE STUDY: OVERBOOKING

- *M Airlines* has a commuter plane capable of flying 16 passengers
- The plane is used on a route for which *M Airlines* charges *$* 225
- The airliner's cost structure is based on

the fixed costs for each flight	*$* 900
the variable costs/passenger	*$* 100
the "no-show" rate	4 %

CASE STUDY – OVERBOOKING

- The refund policy specifies that unused tickets are refunded only if a reservation is cancelled $24\ h$ ahead of the scheduled departure
- The overbooking policy pays $\$\ 100$ as an incentive to each bumped passenger and refunds the ticket
- The decision required is to determine how many reservations should the airliner sell on this plane

SAMPLE CALCULATION FOR THE CASE 18 RESERVATIONS ARE SOLD

total revenues : $\underset{\sim}{R} = 225 \cdot 18 = 4,050$

passenger fixed and variable costs :

$$C_1 = 900 + 100 \cdot min\{number\ of\ "shows", 16\}\ \$$$

bumping costs :

$$C_2 = (225 + 100) \cdot max\{0, number\ of\ "shows" - 16\}\ \$$$

refunds to customers

total costs : $\underset{\sim}{C} = C_1 + C_2$

CASE STUDY: OVERBOOKING

- We evaluate

$$P\left\{ no.\ of\ \text{"}shows\text{"} > 16 \;\middle|\; reservations\ sold = 18 \right\}$$

- We assume that each reservation is a *r.v.* $\underset{\sim}{P}_i$:

$$\underset{\sim}{P}_i = \begin{cases} 1 & passenger\ \ i\ \ is\ a\ \text{"}show\text{"}\ \ with\ \ prob.\ 0.96 \\ \\ 0 & passenger\ \ i\ \ is\ a\ \text{"}no\ show\text{"}\ \ with\ \ prob.\ 0.04 \end{cases}$$

CASE STUDY: OVERBOOKING

- Given reservations sold = 18 , then we need to evaluate

$$P\left\{ \sum_{i=1}^{18} \underset{\sim}{P}_i > 16 \,\middle|\, 18 \; reservations \right\}$$

- We first evaluate

$$P\left\{ \sum_{i=1}^{17} \underset{\sim}{P}_i > 16 \,\middle|\, 17 \; res \right\} = P\left\{ \sum_{i=1}^{17} \underset{\sim}{P}_i \geq 17 \,\middle|\, 17 \; res \right\} = P\left\{ \sum_{i=1}^{17} \underset{\sim}{P}_i = 17 \,\middle|\, 17 \; res \right\}$$

binomial ***r.v.*** **with** $p = 0.96$

CASE STUDY: OVERBOOKING

$$= (.96)^{17}(.04)^{0} \longleftarrow 0.4996$$

- Then,

$$P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i > 16 \middle| 18\, res\right\} = P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i \geq 17 \middle| 18\, res\right\} =$$

$$\underbrace{P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i = 17 \middle| 18\, res\right\}}_{18(.4996)(.04)} + \underbrace{P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i = 18 \middle| 18\, res\right\}}_{(.4996)(.96)} = 0.8359$$

CASE STUDY: OVERBOOKING

- Given reservations sold = 19, then we compute and show that

$$P\left\{\sum_{i=1}^{19} \underset{\sim}{P}_i > 16 \,\middle|\, 19\, res\right\} = .9616$$

- We next consider the profits *r.v.* $\underset{\sim}{\pi}$, where,

$$\underset{\sim}{\pi} = \underset{\sim}{R} - \underset{\sim}{C} = \underset{\sim}{R} - \left(C_1 + C_2\right)$$

and evaluate $E\left\{\underset{\sim}{\pi}\right\}$ for different values of reserva– tions sold

CASE STUDY: OVERBOOKING

- For reservations = 16

$$E\{\underset{\sim}{R}\} = (16)(225) = 3,600$$

$$E\{C_1\} = 900 + 100 \sum_{n=0}^{16} n P\left\{ \sum_{i=1}^{16} \underset{\sim}{P}_i = n \right\}$$

$$= 900 + 100 \; \underbrace{E\left\{ \sum_{n=1}^{16} \underset{\sim}{P}_i \right\}}_{(16)(.96) = 15.36}$$

binomial distribution

$$= 900 + 1,536$$

CASE STUDY: OVERBOOKING

$$= 2,436;$$

also,

$$E\left\{C_2\right\} = \left(225 + 100\right) max \left\{0, \sum_{i=1}^{16} \underset{\sim}{P}_i - 16\right\} = 0$$

and so

$$E\left\{\underset{\sim}{\pi} \middle| 16\ res\right\} = E\left\{\underset{\sim}{R}\right\} - E\left\{\underset{\sim}{C}\right\}$$

$$= E\left\{\underset{\sim}{R}\right\} - E\left\{C_1 + C_2\right\}$$

$$= 3,600 - 2,436$$

$$= 1,164$$

CASE STUDY: OVERBOOKING

- For reservations = 17

$$E\{\underset{\sim}{R}\} = (17)(225) = 3,825$$

$$E\{C_1\} = 900 + 100\sum_{n=0}^{16} nP\left\{\sum_{i=1}^{17} \underset{\sim}{P}_i = n\right\} + 100.16 \bullet P\left\{\sum_{i=1}^{17} \underset{\sim}{P}_i = 17\right\}$$

$$= 900 + 782.70 + 799.34$$

$$= 2,482.04$$

CASE STUDY: OVERBOOKING

also,

$$E\left\{C_2\right\} = 325P\left\{\sum_{i=1}^{17} \underset{\sim}{P}_i = 17\right\}$$

$$= 325\left(0.4996\right)$$

$$= 162.37$$

and so

$$E\left\{\underset{\sim}{\pi}\middle| 17\ res\right\} = 3,825 - 2,482.04 - 162.37$$

$$= 1,180.59 > 1,164$$

CASE STUDY: OVERBOOKING

- **For reservations = 18**

$$E\{\underset{\sim}{R}\} = (18)(225) = 4{,}050$$

$$E\{C_1\} = 900 + 100\sum_{n=0}^{16} nP\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i = n\right\} + 1{,}600 \cdot P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i > 16\right\}$$

$$= 900 + 253.22 + 1{,}342.89$$

$$= 2{,}496.11$$

CASE STUDY: OVERBOOKING

$$E\left\{C_2\right\} = 325\,\underbrace{P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i = 17\right\}}_{.3597} + 650\,\underbrace{P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i = 18\right\}}_{.4796} = 428.65$$

and

$$E\left\{\underset{\sim}{\pi} \,\middle|\, 18\ res\right\} = 4{,}050 - 2{,}496.11 - 428.65$$

$$= 1{,}125.24$$

$$< 1{,}180.59$$

CASE STUDY: OVERBOOKING

- We can show that for reservations = 19

$$E\left\{ \underset{\sim}{\pi} \;\middle|\; 19\ res \right\} < 1180.59$$

- We conclude that the profits are maximized for reservations = 17 and so any overbooking over that number results in lower profits

