ECE 307 – Techniques for Engineering Decisions

Probability Distributions

George Gross

Department of Electrical and Computer Engineering

University of Illinois at Urbana-Champaign

OUTLINE OF DISTRIBUTION REVIEWED

- Discrete
  - Binomial
  - Poisson
- Continuous
  - 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, correct/incorrect, on/off or 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:

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

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

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

BINOMIAL DISTRIBUTION EXAMPLE

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

- **The probability that 3 or more machines are on is determined by evaluating**

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

BINOMIAL DISTRIBUTION EXAMPLE

$$= P\{3 \text{ machines are on}\} +$$

$$P\{4 \text{ machines are on}\} +$$

$$\cdots \quad +$$

$$P\{8 \text{ machines are on}\}$$

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

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

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

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

the binomial distribution

THE BINOMIAL DISTRIBUTION

- We can show that:

$$E\left\{\underset{\sim}{R}\right\} = np$$

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

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

EXAMPLE: SOFT PRETZELS

- **Pretzel entrepreneur can sell pretzels at *$ 0.50 per unit* with a market potential of 100,000 pretzels within a year; there exists a competing product and so we know he cannot sell that many**
- **Basic model is binomial:**

new pretzel is a hit (success)	$\Rightarrow$	**captures 30% of market in one year**
new pretzel is a flop (flop)	$\Rightarrow$	**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 being taste tested against the competition; the result is that 5 out of 20 people prefer the new pretzel
- We evaluate the conditional probability

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

EXAMPLE: SOFT PRETZELS

- We define the success *r.v.*

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

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

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

0.179 from the binomial table

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

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

0.0032 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,

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

$$= \frac{(0.179)(0.5)}{(0.179)(0.5) + (0.032)(0.5)}$$

$$= 0.848$$

THE POISSON DISTRIBUTION

- The binomial distribution is good for representing successes in repeated trials
- The Poisson distribution is appropriate for representing specific events over time or space: e.g., number of customers who are served by a butcher in a meat market, or 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 so events do not happen *too frequently***

REQUIREMENTS FOR A POISSON DISTRIBUTION

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

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

is fixed and *independent* of all other events

- **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

$$P\{\underset{\sim}{X} = k\} = \frac{e^{-m} m^k}{k!}$$

$$E\{\underset{\sim}{X}\} = m \qquad var\{\underset{\sim}{X}\} = m$$

m 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
  - 1024 units are produced
  - based on past experience, a flawed product is manufactured every 197 units and so, on average, there are that $\frac{1024}{197} \approx 5.2$ flawed units in the 1024 products are produced

EXAMPLE: POISSON DISTRIBUTION

- Note that the Poisson conditions are satisfied
  - the sample has 1024 units
  - there are only a few flawed units in the 1024 sample
  - the probability of a flawed unit is small
  - each flawed unit is *independent* of every other flawed unit

EXAMPLE: POISSON DISTRIBUTION

- **Poisson distribution is appropriate represen–tation with $m = 5.2$ and so,**

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

- **If we want to determine the probability of 4 or more flawed units, we compute**

EXAMPLE: POISSON DISTRIBUTION

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

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

- **The Poisson table states that**

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

cuu duong than cong. com

and therefore

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

EXAMPLE: SOFT PRETZELS

- The pretzel enterprise is going well: several retail outlets and a street vendor sell 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 considered to be

  "good" if 20 *p/h* are sold with probability 0.7

  "bad" if 10 *p/h* are sold with probability 0.2

  "dismal" if 6 *p/h* are sold with probability 0.1

EXAMPLE: SOFT PRETZELS

- After the first week, long enough to make a mark, a 30 – minute test is run and 7 pretzels are sold during the 30 – minute test period

- We define

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

and assume Poisson distribution applies

EXAMPLE: SOFT PRETZELS

- We determine the conditional probabilities of the new location conditioned on the test outcomes

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

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

- We compute

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

EXAMPLE: SOFT PRETZELS

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

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

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

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

EXAMPLE: SOFT PRETZELS

$$P\left\{\underset{\sim}{L}=\text{"good"} \,\middle|\, \underset{\sim}{X}=7\right\} = \frac{(0.09)(0.7)}{(0.09)(0.7)+(0.104)(0.2)+(0.022)(0.1)}$$

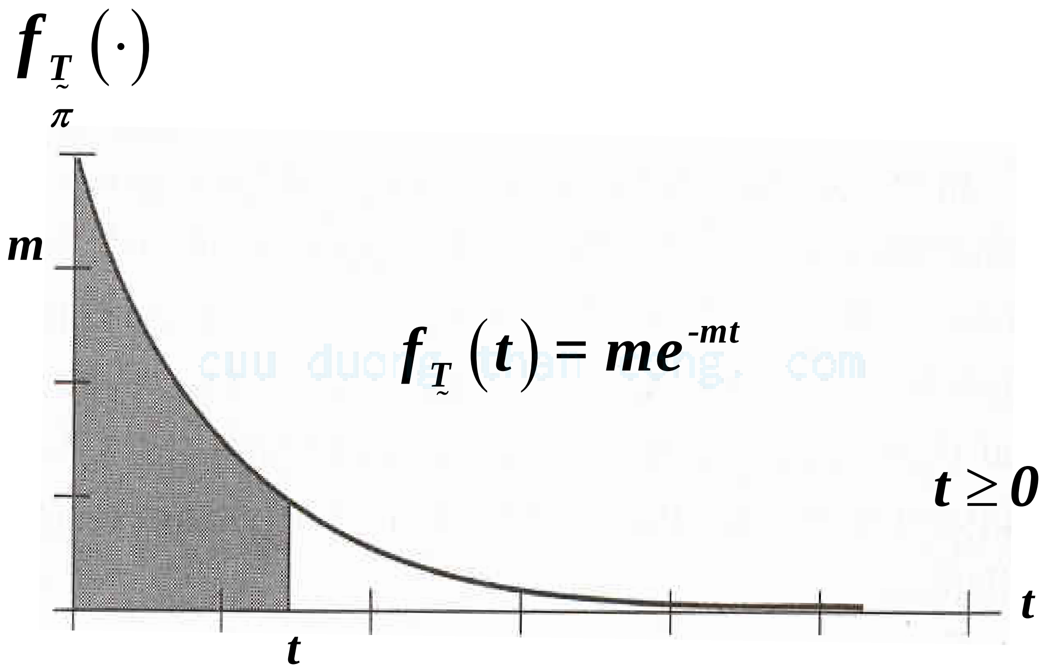
$$= 0.733$$

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

$$P\left\{\underset{\sim}{L}=\text{"dismal"} \,\middle|\, \underset{\sim}{X}=7\right\} = 0.025$$

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
- Consider the Poisson distributed *r.v.* $\underset{\sim}{X}$ with representing the number of events in a given quantity of measure, e.g., period of time
- We define $\underset{\sim}{T}$ to be the *r.v.* for the uncertain quantity of measure, e.g., time between two sequential events

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 m parameter**

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** $\longleftrightarrow$ $m = 20$ *pretzels / hour*

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

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

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

EXAMPLE: SOFT PRETZELS

$$P\{\underset{\sim}{T} > 3.5\, minutes\} = P\{\underset{\sim}{T} > 3.5\, minutes \mid m = 20\}\; P\{m = 20\} +$$

$$P\{\underset{\sim}{T} > 3.5\, minutes \mid m = 10\}\; P\{m = 10\} +$$

$$P\{\underset{\sim}{T} > 3.5\, minutes \mid m = 6\}\; P\{m = 6\}$$

$\equiv 0.0583\, hour$

- We evaluate

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

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

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

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

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

EXAMPLE: SOFT PRETZELS

and so

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

- Therefore,

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

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 are governed by the *normal* distribution, or commonly known as the *bell curve***

cuu duong than cong. com

- **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

mean

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

variance

standard deviation

with

$$E\{\underset{\sim}{Y}\} = \mu \qquad \textit{and} \qquad var\{\underset{\sim}{Y}\} = \sigma^2$$

THE STANDARD NORMAL DISTRIBUTION

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

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

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

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

with

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

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, all the values of the normal distribution can be obtained from the *standard normal distribution* through 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 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 4*mm,* but in reality, the width is a normally distributed *r.v.* $\underset{\sim}{W}$ with

EXAMPLE: QUALITY CONTROL

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

and

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

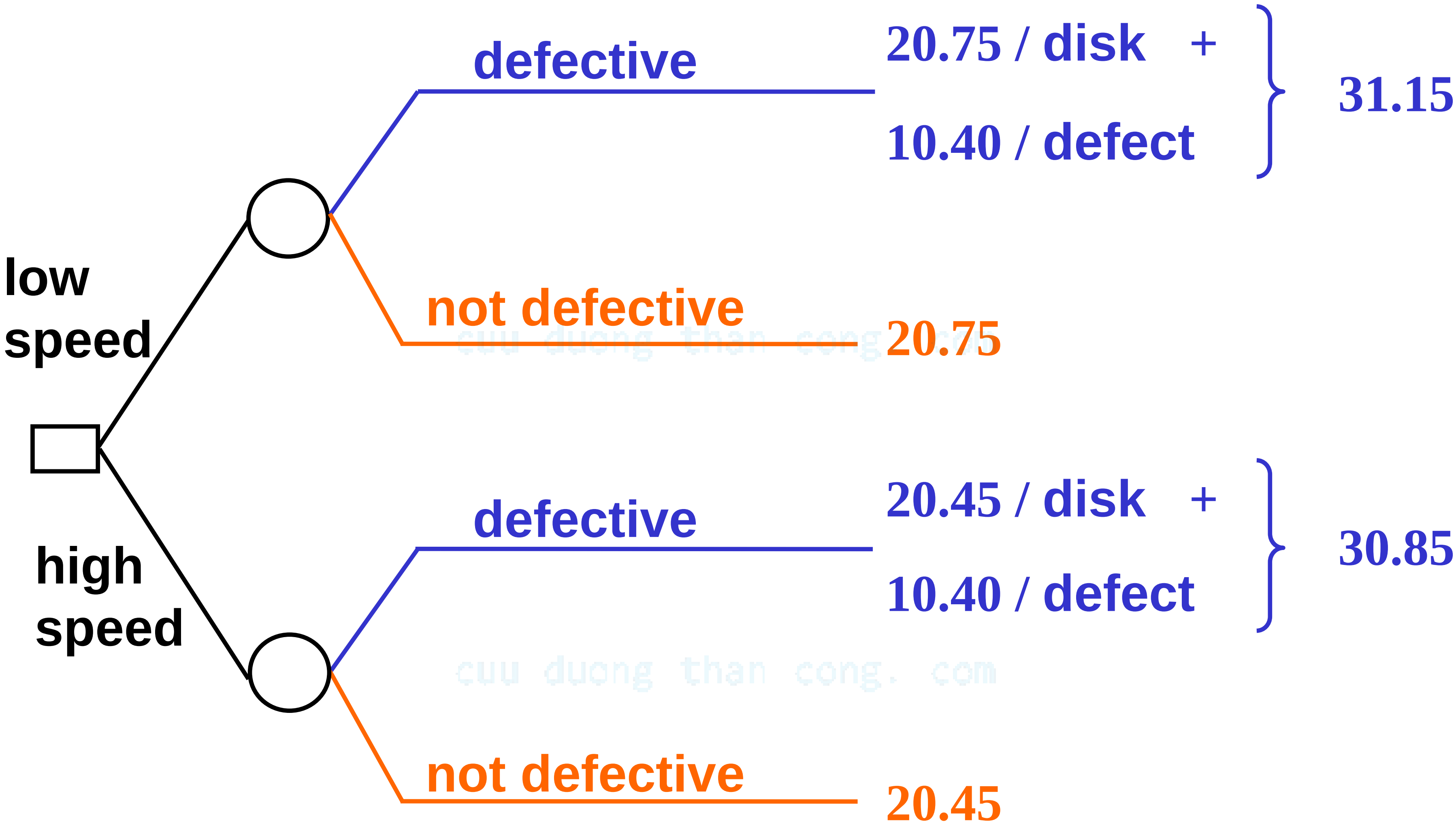
- **The corresponding costs (*$*) of the disk drive are**

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 lesser costs at the high speed
- We need 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 need to evaluate the probability of each outcome

LOW – SPEED PROBABILITY EVALUATION

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

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

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

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

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

LOW – SPEED PROBABILITY EVALUATION

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

$$1 - \Big[\underbrace{P\left\{\underset{\sim}{Z} \leq 2.63\right\}}_{0.9957} - \underbrace{P\left\{\underset{\sim}{Z} \leq -2.63\right\}}_{0.0043}\Big] = 0.0086$$

$$\underbrace{0.9957 \qquad 0.0043}_{0.9914}$$

HIGH – SPEED PROBABILITY EVALUATION

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

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

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

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

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

HIGH – SPEED PROBABILITY EVALUATION

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

$$1 - \Big[\underbrace{\underbrace{P\left\{\underset{\sim}{Z} \le 1.92\right\}}_{0.9726} - \underbrace{P\left\{\underset{\sim}{Z} \le -1.92\right\}}_{0.0274}}_{0.9452}\Big] = 0.0548$$

MEAN VALUE EVALUATION

- We next evaluate the mean cost per disk

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

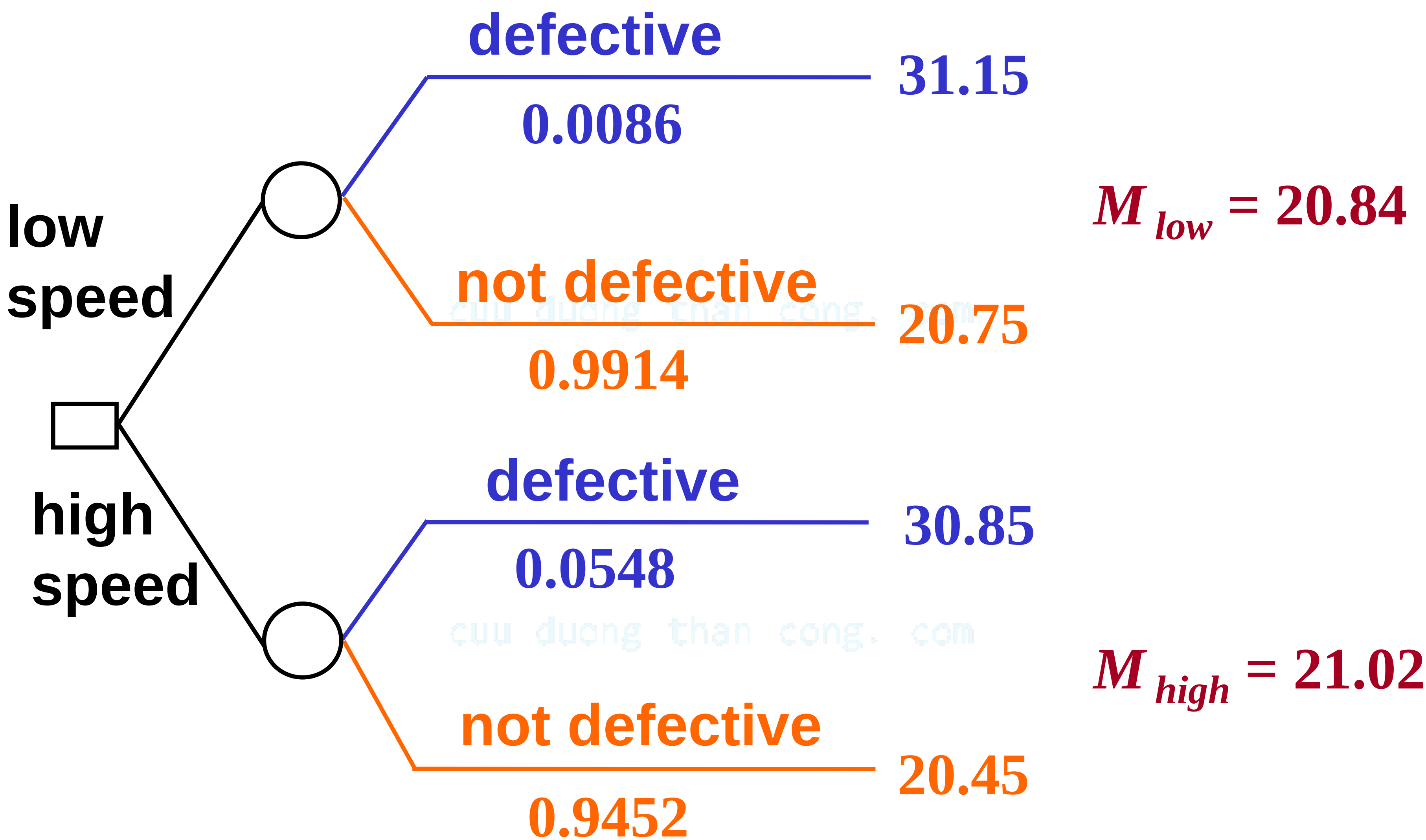
$$= 20.84$$

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

$$= 21.02$$

- We summarize the information in the decision tree

EXAMPLE: QUALITY CONTROL



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

SAMPLE CALCULATION FOR SELLING 18 RESERVATIONS

total revenues : $R = 225 * 18 = 4050$

passenger fixed and variable costs :

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

bumping costs :

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

refunds to customers

total costs : $C = C_1 + C_2$

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

- If 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

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

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

CASE STUDY: OVERBOOKING

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

- Then,

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

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

$$= 0.8359$$

CASE STUDY: OVERBOOKING

- If reservations sold $= 19$, then we can compute and show that

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

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

$$\underset{\sim}{\pi} = \underset{\sim}{R} - \underset{\sim}{C} = \underset{\sim}{R} - \left(\underset{\sim}{C}_1 + \underset{\sim}{C}_2\right)$$

and evaluate $E\{\underset{\sim}{\pi}\}$ for different values of reser–vations sold

CASE STUDY: OVERBOOKING

- For reservations = 16

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

$$E\{\underset{\sim}{C}_1\} = 900 + 100 \sum_{n=0}^{16} nP\left\{\sum_{i=1}^{16} \underset{\sim}{P}_i = n\right\}$$

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

binomial distribution

$$= 900 + 1536$$

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

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

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

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

$$= 1{,}164$$

CASE STUDY: OVERBOOKING

- For reservations = 17

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

$$E\{\underset{\sim}{C}_1\} = 900 + 100\sum_{n=0}^{16} nP\left\{\sum_{i=1}^{17}\underset{\sim}{P}_i = n\right\} + 100.16 \cdot P\left\{\sum_{i=1}^{17}\underset{\sim}{P}_i = 17\right\}$$

$$= 900 + 782.70 + 799.34$$

$$= 2,482.04$$

CASE STUDY: OVERBOOKING

also,

$$E\left\{\underset{\sim}{C}_2\right\} = 325 P\left\{\sum_{i=1}^{17} \underset{\sim}{P}_i = 17\right\}$$

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

$$= 162.37$$

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

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

$$E\left\{\underset{\sim}{C}_1\right\} = 900 + 100\sum_{n=0}^{16} nP\left\{\sum_{i=1}^{18}\underset{\sim}{P}_i = n\right\} + 1,600 \cdot P\left\{\sum_{i=1}^{18}\underset{\sim}{P}_i > 16\right\}$$

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

$$= 2,496.11$$

CASE STUDY: OVERBOOKING

$$E\left\{\underset{\sim}{C}_2\right\} = 325\,\underbrace{P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i = 17\right\}}_{.3597} + 650\,\underbrace{P\left\{\sum_{i=1}^{18} \underset{\sim}{P}_i = 18\right\}}_{.4796} = 428.65$$

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 if reservations = 17 and so any greater overbooking is at the sacrifice of lower profits

CASE: MUNI SOLID WASTE

- This case concerns the "risks" posed by constructing an incinerator for disposal of a city's solid waste
- There is no question of "whether" to construct an incinerator since landfill was to be full within 3 years and no other choices are apparent

CASE: MUNI SOLID WASTE

- **Of particular interest are the *residual* emissions that need to be *estimated* for**
  - ***dioxins***
  - ***furans* (organic compounds)**
  - ***particulate matter***
  - **SO_2 (represents acid gases)**

CASE: MUNI SOLID WASTE

- The specifications for a 250 *ton/day* incinerator – borderline between the EPA – classified small and medium plants – have to meet the EPA's proposed emission levels for the three key pollutants

  - *dioxins/furans* – denoted by $\underset{\sim}{D}$
  - *particulate matter* – denoted by $\underset{\sim}{PM}$
  - *sulphur dioxide* – denoted by $\underset{\sim}{SO_2}$

CASE: MUNI SOLID WASTE

pollutant	plant capacity in *tons of waste per day*	
	small (*below* 250)	medium (*above* 25)
dioxins/furans mg/nm^3	500	125
$\underset{\sim}{PM}$ *(mg/dscm)*	69	69
$\underset{\sim}{SO}_2$ *(ppmdv)*	–	30

CASE: MUNI SOLID WASTE

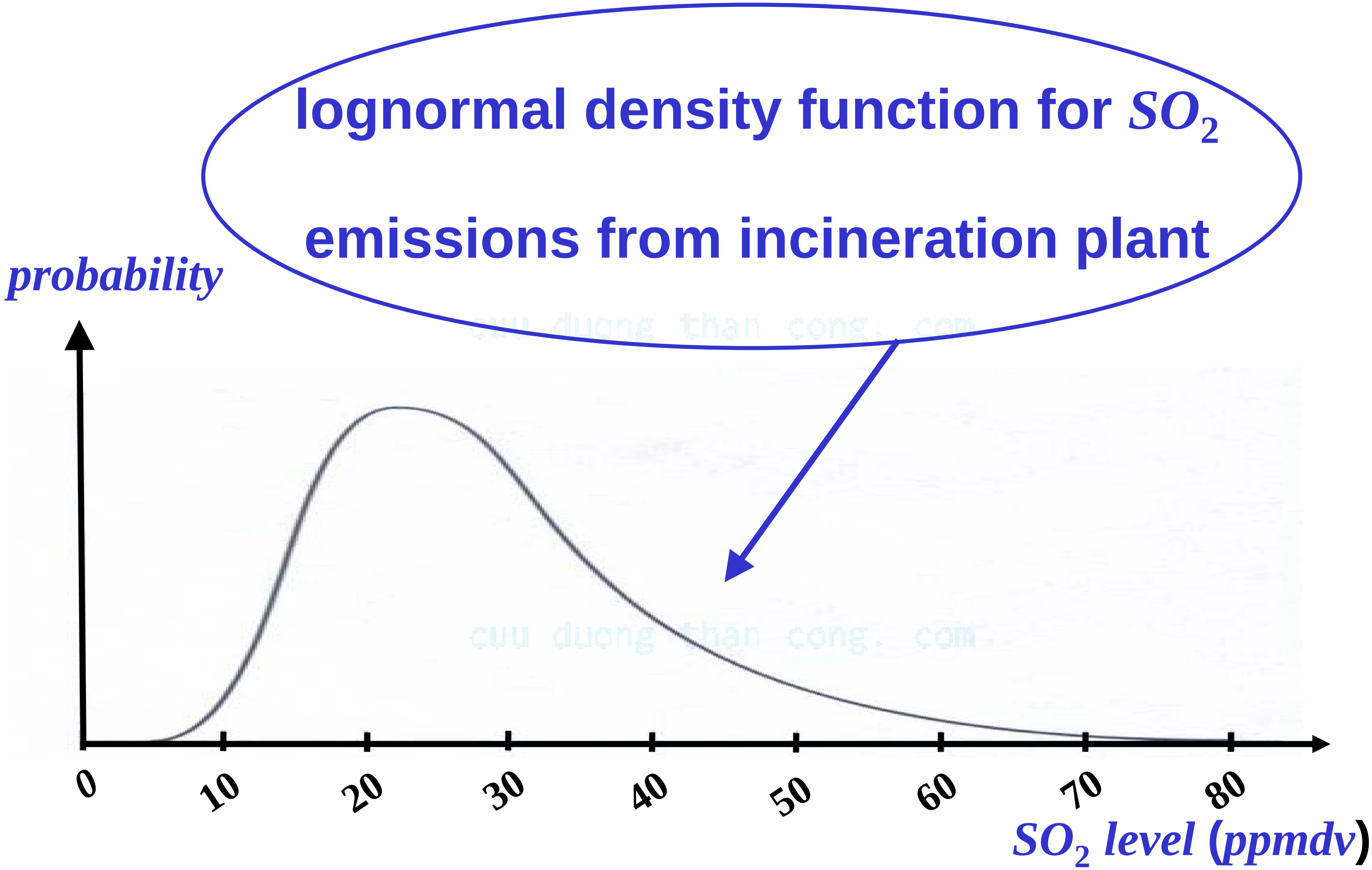
- The typical approach in environmental risk analysis is a "worst case" scenario assessment which fails to capture the *uncertainty* present in both the amount of waste and the contents of the waste

CASE: MUNI SOLID WASTE

- The lognormal distribution is used to represent the distribution for emission levels
- Lognormal distribution parameters μ and σ for pollutants

pollutant r.v.	μ	σ
$\underset{\sim}{D}$	3.13	1.20
$\underset{\sim}{PM}$	3.43	0.44
$\underset{\sim}{SO_2}$	3.20	0.39

CASE: MUNI SOLID WASTE



CASE: MUNI SOLID WASTE

- We define the *r.v.* $\underset{\sim}{X}$, which is lognormally distributed with parameters μ and σ, by

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

$$E\left\{\underset{\sim}{X}\right\} = e^{\left(\mu+\frac{1}{2}\sigma^{2}\right)} \quad \textbf{and} \quad var\left\{\underset{\sim}{X}\right\} = e^{2\mu}\left(e^{\sigma^{2}}-1\right)e^{\sigma^{2}}$$

- We evaluate the probability of exceeding the small plant levels of emissions

CASE: MUNI SOLID WASTE

$$P\{\underset{\sim}{D} > 500\} = P\{\underset{\sim}{Y} = ln\,\underset{\sim}{D} > ln(500)\}$$

$$\underset{\sim}{Z} = \frac{\underset{\sim}{Y} - 3.13}{1.2}$$

affine transformation

$$= P\left\{\underset{\sim}{Z} > \frac{ln(500) - 3.13}{1.2}\right\}$$

$$= P\{\underset{\sim}{Z} > 2.57\} \longleftarrow 0.0051$$

CASE: MUNI SOLID WASTE

$$P\left\{\underset{\sim}{PM} > 69\right\} = P\left\{\underset{\sim}{Y} = ln\left(\underset{\sim}{PM}\right) > ln\left(69\right)\right\}$$

$$\underset{\sim}{Z} = \frac{\underset{\sim}{Y} - 3.43}{0.44}$$

affine transformation

$$= P\left\{\underset{\sim}{Z} > \frac{ln\left(69\right) - 3.43}{0.44}\right\}$$

$$= P\left\{\underset{\sim}{Z} > 1.83\right\}$$

0.0336

CASE: MUNI SOLID WASTE

$$P\left\{\underset{\sim}{SO}_2 > 30\right\} = P\left\{\underset{\sim}{Y} = ln\left\{\underset{\sim}{SO}_2\right\} > ln\left(30\right)\right\}$$

$$\underset{\sim}{Z} = \frac{\underset{\sim}{Y} - 3.20}{0.39}$$

affine transformation

$$= P\left\{\underset{\sim}{Z} > \frac{ln\left(30\right) - 3.20}{0.39}\right\}$$

0.3015

$$= P\left\{\underset{\sim}{Z} > 0.52\right\}$$

the probability of a single observation

CASE: MUNI SOLID WASTE

- In practice, $\underset{\sim}{SO}_2$ has to be monitored continuously and the average daily emission level *must remain* below the level specified in the table
- If we take 24 hourly observations $\underset{\sim}{X}_i$ of $\underset{\sim}{SO}_2$ levels and define the geometric mean

$$\underset{\sim}{G} = \left[\prod_{i=1}^{n} \underset{\sim}{X}_i \right]^{\frac{1}{n}}, \quad n = 24$$

CASE: MUNI SOLID WASTE

then, we can show that $\underset{\sim}{G}$ is also lognormal with parameters $\left(\mu , \dfrac{\sigma}{\sqrt{24}} \right)$

$$P\left\{\underset{\sim}{G} > 30\right\} = P\left\{\underset{\sim}{Y} = ln\left(\underset{\sim}{G}\right) > ln\left(30\right)\right\}$$

$$\underset{\sim}{Z} = \frac{\underset{\sim}{Y} - 3.20}{0.39/\sqrt{24}}$$

affine transformation

$$= \underbrace{P\left\{\frac{\underset{\sim}{Z} - 3.2}{0.39/\sqrt{24}} > 2.53\right\}}_{0.0057}$$

CASE: MUNI SOLID WASTE

- Note that this is a much smaller probability than for a single observation and leads to a more realistic assessment of the probability
- The requirement of a small plant are therefore met

