ECE 307 – Techniques for Engineering Decisions

11. Basic Probability: Case Studies

George Gross

Department of Electrical and Computer Engineering

University of Illinois at Urbana-Champaign

OIL WILDCATTING: SITE DATA

- We consider two possible exploratory well sites
  - site 1: rather uncertain
  - site 2: fairly certain for a low production level
- Geological fact: if the rock strata underlying site 1 are characterized by a "dome" structure, the chances are better to find oil than if "no dome" structure exists

OIL WILDCATTING: SITE DATA

state	*site 1 with $ 100k drilling costs*	*site 2 with $ 200k drilling costs*	
	payoffs (k$)	*probability*	*payoffs (k$)*
dry	– 100	0.2	– 200
low production	150	0.8	50
high production	500	*0*	–

MODELING OF SITE 1 UNCERTAINTY

$$\underset{\sim}{S} = \textit{structure r.v.} = \begin{cases} \textit{dome structure} & \textit{with prob } 0.6 \\ \textit{other} & \textit{with prob } 0.4 \end{cases}$$

conditioning on the event $\left\{ \underset{\sim}{S} = \textit{dome} \right\}$

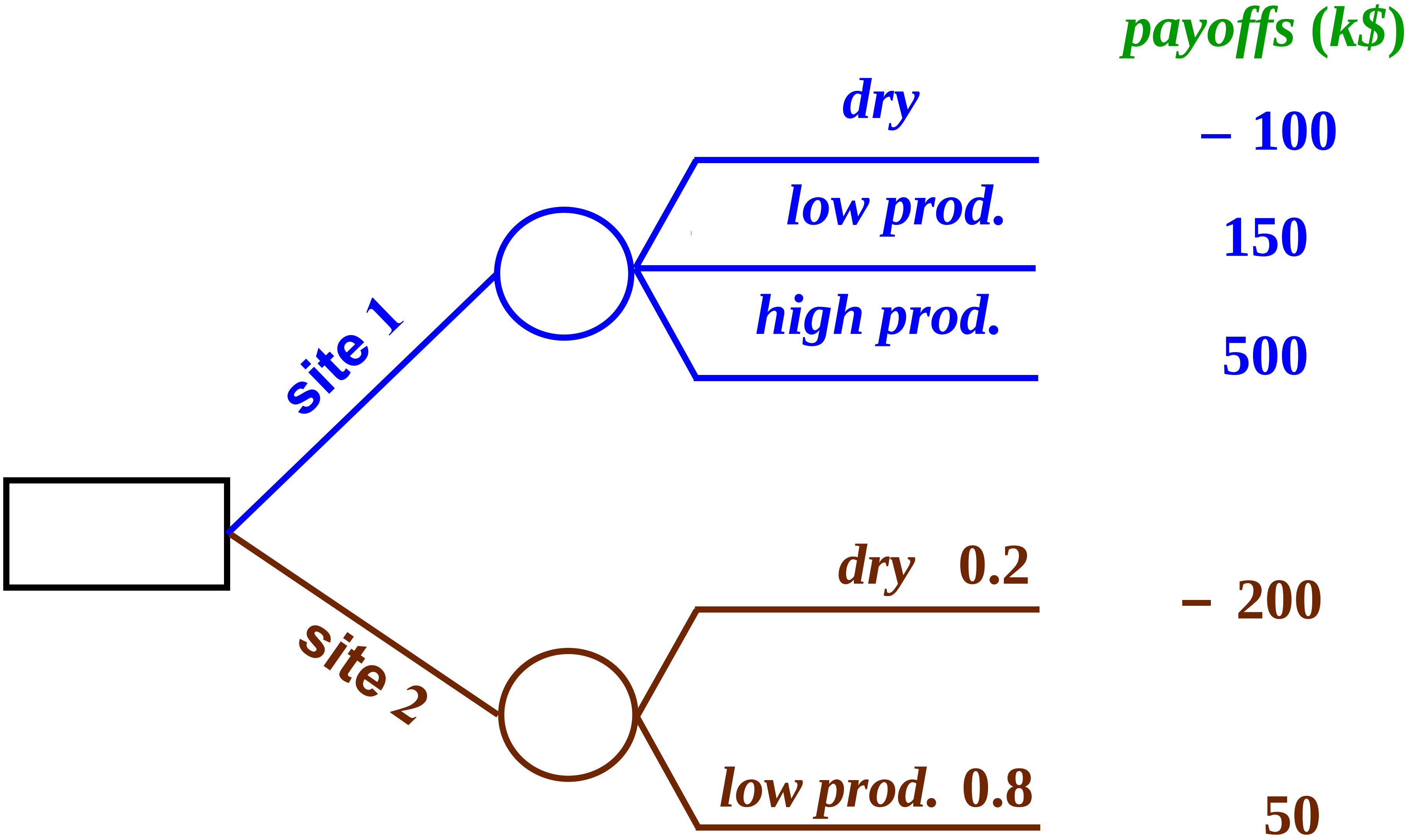
state $\underset{\sim}{X}$ *outcome*	$P\left\{ \textit{state } \underset{\sim}{X} = x \,\middle|\, \underset{\sim}{S} = \textit{dome} \right\}$
dry	**0.60**
low production	**0.25**
high production	**0.15**

SITE 1: NO DOME OUTCOMES

conditioning on the event $\left\{ \underset{\sim}{S} = \textit{no dome} \right\}$

state outcome x	$P\left\{ \textit{state } \underset{\sim}{X} = x \,\middle|\, \underset{\sim}{S} = \textit{no dome} \right\}$
dry	**0.850**
low production	**0.125**
high production	**0.025**

DECISION TREE CONSTRUCTION



COMPUTATION OF PROBABILITIES OF STATES FOR SITE 1

$$P\{dry\} = P\{state\ of\ site\ 1 = dry\}$$

$$= P\{state = dry \mid \underset{\sim}{S} = dome\} \cdot P\{\underset{\sim}{S} = dome\} +$$

$$P\{state = dry \mid \underset{\sim}{S} = no\ dome\} \cdot P\{\underset{\sim}{S} = no\ dome\}$$

$$= (0.6)(0.6) + (0.85)(0.4)$$

$$= 0.7$$

COMPUTATION OF PROBABILITIES OF STATES FOR SITE 1

$$P\{low\ \ prod.\} = P\{state\ of\ site\ 1 = low\ \ prod.\}$$

$$= P\{state = low\ prod. \mid S = dome\} \cdot P\{S = dome\} +$$

$$P\{state = low\ prod. \mid S = no\ dome\} \cdot P\{S = no\ dome\}$$

$$= (0.25)(0.6) + (0.125)(0.4)$$

$$= 0.2$$

CONFIGURATION OF PROBABILITIES OF STATES FOR SITE 1

$$P\left\{high\ \ prod.\right\} = P\left\{state\ \underset{\sim}{X}\ of\ site\ 1 = high\ \ prod.\right\}$$

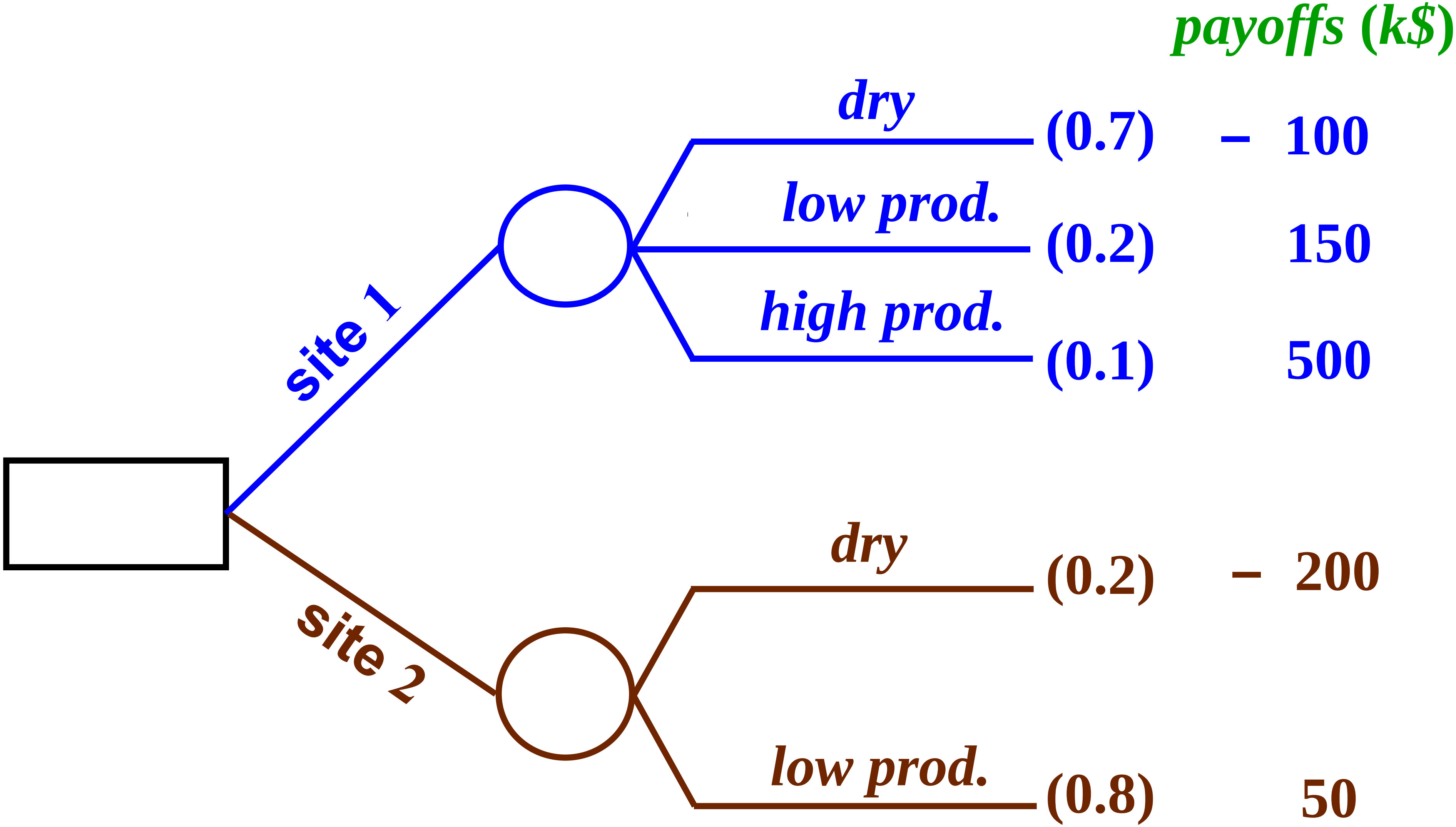
$$= P\left\{state\ \underset{\sim}{X} = high\ \ prod.\ \middle|\ \underset{\sim}{S} = dome\right\} \cdot P\{\underset{\sim}{S} = dome\} +$$

$$P\left\{state\ \underset{\sim}{X} = high\ \ prod.\ \middle|\ \underset{\sim}{S} = no\ \ dome\right\} \cdot P\{\underset{\sim}{S} = no\ \ dome\}$$

$$= (0.15)(0.6) + (0.025)(0.4)$$

$$= 0.1$$

DECISION DIAGRAM COMPLETION



EVALUATION OF PAYOFFS

- Site 1 evaluation:

$$\underbrace{E\{payoffs\}}_{EMV} = \sum (payoffs \text{ in state } \underset{\sim}{X} = x)\, P\{state\ \underset{\sim}{X} = x\}$$

$$= -100 \cdot (0.7) + 150 \cdot (0.2) + 500 \cdot (0.1)$$

$$= 10\, k\$$$

- Site 2 evaluation:

$$E\{payoffs\} = -200 \cdot (0.2) + 50 \cdot (0.8)$$

$$= 0\, k\$$$

VARIANCE EVALUATION

- Site 1 evaluation:

$$\sigma_1^2 = 0.7\left[-100-10\right]^2 + 0.2\left[150-10\right]^2 + 0.1\left[500-10\right]^2$$

$$= 36,400\left(k\$\right)^2$$

and so

$$\sigma_1 = 190.8\ k\$$$

- Site 2 evaluation:

$$\sigma_2^2 = 0.2\left[-200-0\right]^2 + 0.8\left[50-0\right]^2$$

$$= 10,000\left(k\$\right)^2$$

VARIANCE EVALUATION

and so

$$\sigma_2 = 100k\$$$

- Therefore site 1 has greater variability and

  therefore greater perceived *risk* than site 2 since

$$\sigma_1 \approx 2\sigma_2 > \sigma_2$$

PROBABILITY EVALUATION

state outcome x	$P\{state = x\}$	$P\{\underset{\sim}{X} = x \mid \underset{\sim}{S} = s\}\, P\{\underset{\sim}{S} = s\}$	
		$s = dome$	$s = no\ dome$
dry	0.7	0.36	0.34
low prod.	0.2	0.15	0.05
high prod.	0.1	0.09	0.01
$P\{\underset{\sim}{S} = s\}$		0.60	0.40

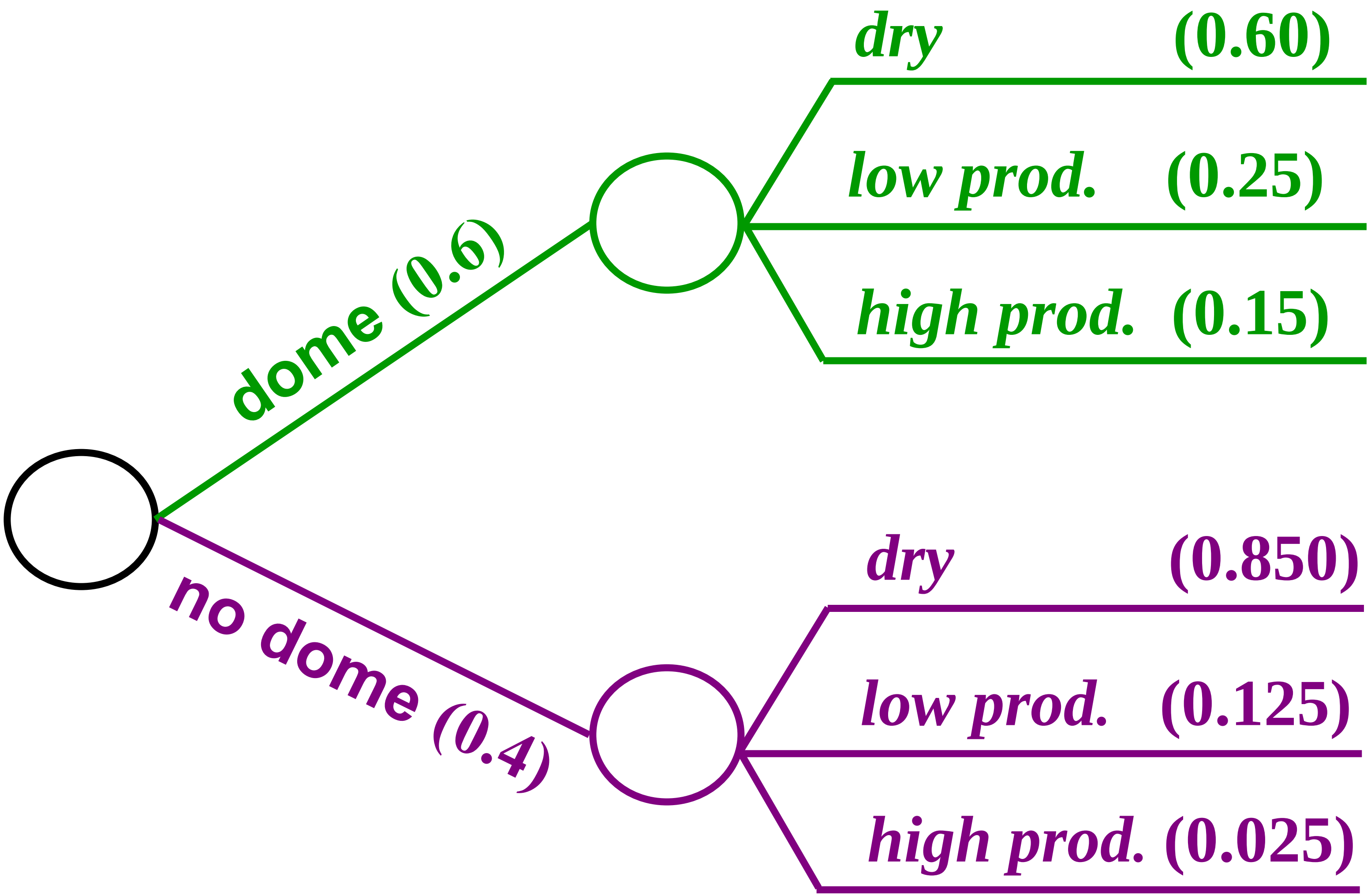
JOINT PROBABILITIES

$$P\{\underset{\sim}{state} = low\ prod\ \ and\ \underset{\sim}{S} = dome\}$$

$$= \underbrace{P\left\{\underset{\sim}{state} = low\ \ prod \,\middle|\, \underset{\sim}{S} = dome\right\}}_{\mathbf{0.25}} \cdot \underbrace{P\left\{\underset{\sim}{S} = dome\right\}}_{\mathbf{0.6}}$$

$$= \mathbf{0.15}$$

DECISION DIAGRAM WITH PROBABILITIES



REVERSE PROBABILITIES

$$P\left\{\underset{\sim}{S} = dome \,\middle|\, \underset{\sim}{state} = dry\right\}$$

$$= \frac{P\left\{\underset{\sim}{S} = dome \text{ and } \underset{\sim}{state} = dry\right\}}{P\left\{\underset{\sim}{state} = dry\right\}}$$

$$= \frac{P\left\{\underset{\sim}{state} = dry \,\middle|\, \underset{\sim}{S} = dome\right\} \cdot P\left\{\underset{\sim}{S} = dome\right\}}{P\left\{\underset{\sim}{state} = dry\right\}}$$

$$P\{\underset{\sim}{state} = dry\} = P\{\underset{\sim}{state} = dry \mid \underset{\sim}{S} = dome\} \cdot P\{\underset{\sim}{S} = dome\} +$$

$$P\{\underset{\sim}{state} = dry \mid \underset{\sim}{S} = no\ dome\} \cdot P\{\underset{\sim}{S} = no\ dome\}$$

REVERSE PROBABILITIES

$$P\left\{\underset{\sim}{S} = dome \,\middle|\, \underset{\sim}{state} = dry\right\} = \frac{(0.6)(0.6)}{(0.6)(0.6) + (0.85)(0.4)}$$

$$= \frac{0.36}{0.36 + (0.85)(0.4)}$$

$$= \frac{0.36}{0.70}$$

$$= 0.51$$

$$P\{\underset{\sim}{S} = no\ dome \mid \underset{\sim}{state} = dry\} = 1 - P\{\underset{\sim}{S} = dome \mid \underset{\sim}{state} = dry\}$$

$$= 1 - 0.51$$

$$= 0.49$$

DECISION ANALYSIS MONTHLY PROBLEM: MAY DATA

May subscription data	*expiring subscriptions (%)*	*renewal ratio (%)*
gift subscriptions	**70**	**75**
promotional subscriptions	**20**	**50**
previous subscribers	**10**	**10**
total	**100**	

DECISION ANALYSIS MONTHLY PROBLEM: JUNE DATA

June subscription data	*expiring subscriptions (%)*	*renewal ratio (%)*
gift subscriptions	**45**	**85**
promotional subscriptions	**10**	**60**
previous subscribers	**45**	**20**
total	**100**	

DECISION ANALYSIS MONTHLY PROBLEM: SUBSCRIPTIONS DATA

- **The concern is that overall proportion of renewals had dropped from May to June**
- **Yet, the table figures indicate that the proportion of renewals had increased in each category**
- **We need to analyze the data in a meaningful fashion and correctly interpret it**

DECISION ANALYSIS MONTHLY PROBLEM

- We can view the data in the two tables as providing probabilities for the renewal *r.v.*

$$\underset{\sim}{R} = \begin{cases} \textit{renewal} \\ \textit{no renewal} \end{cases}$$

- However, the information is given as conditional probabilities with the conditioning on the subscription type with *r.v.* $\underset{\sim}{S}$

$$\underset{\sim}{S} = \begin{cases} \textit{gift} \\ \textit{promotional} \\ \textit{previous} \end{cases}$$

DECISION ANALYSIS MONTHLY PROBLEM

- **We use the May and June data and compute:**

$$P\left\{\underset{\sim}{R} = renewal\right\} = P\left\{\underset{\sim}{R} = renewal \,\middle|\, \underset{\sim}{S} = gift\right\} \cdot P\left\{\underset{\sim}{S} = gift\right\} +$$

$$P\left\{\underset{\sim}{R} = renewal \,\middle|\, \underset{\sim}{S} = promo\right\} \cdot P\left\{\underset{\sim}{S} = promo\right\} +$$

$$P\left\{\underset{\sim}{R} = renewal \,\middle|\, \underset{\sim}{S} = previous\right\} \cdot P\left\{\underset{\sim}{S} = previous\right\}$$

- **The renewal probabilities are computed for each month**

DECISION ANALYSIS MONTHLY PROBLEM

$$P\{\underset{\sim}{R}_{May} = renewal\} = (0.75)(0.7) + (0.5)(0.2) + (0.1)(0.1)$$

$$= 0.635$$

$$P\{\underset{\sim}{R}_{June} = renewal\} = (0.85)(0.45) + (0.6)(0.1) + (0.2)(0.45)$$

$$= 0.5325$$

- Due to the change of the mix,

$$P\{\underset{\sim}{R}_{June} = renewal\} < P\{\underset{\sim}{R}_{May} = renewal\}$$

even though the renewal proportion increased in each category

DISCRIMINATION CASE STUDY

- We explore the relationship between the race of convicted defendants in murder trials and the imposition of the death penalty in these trials on the defendants

- This is a good example to illustrate the care required to correctly interpret the data

DISCRIMINATION CASE STUDY: DATA

defendants		death penalty imposed		total defendants
		yes	no	
race	white	19	141	160
	black	17	149	166
total		36	290	326

DISCRIMINATION CASE STUDY: USING THE DATA

- We define the *r.v.s*

$$\underset{\sim}{D} = death\ \ penalty = \begin{cases} 1 & \text{death penalty is imposed} \\ 0 & \text{otherwise} \end{cases}$$

$$\underset{\sim}{R} = race = \begin{cases} white & \text{defendant is white} \\ black & \text{defendant is black} \end{cases}$$

- We use data of the table to determine

$$P\left\{ \underset{\sim}{D} = 1 \;\middle|\; \underset{\sim}{R} = white \right\} \text{ and } P\left\{ \underset{\sim}{D} = 1 \;\middle|\; \underset{\sim}{R} = black \right\}$$

DISCRIMINATION CASE STUDY: USING THE DATA

- The table provides values

$$P\left\{\underset{\sim}{D}=1 \middle| \underset{\sim}{R}=white\right\} = \frac{19}{160} = 0.119$$

$$P\left\{\underset{\sim}{D}=1 \middle| \underset{\sim}{R}=black\right\} = \frac{17}{166} = 0.102$$

- These two probabilities indicate little difference between the treatment of the two races

- We use additional data to probe a little deeper

DISCRIMINATION CASE STUDY: USING MORE DATA

race of victim	*race of defendant*	*death penalty imposed*		*total defendants*
		yes	*no*	
white	*white*	19	132	151
	black	11	52	63
	total	30	184	214
black	*white*	*0*	9	9
	black	6	97	103
	total	6	106	112
total for all cases		36	290	326

DISCRIMINATION CASE STUDY: USING MORE DATA

- Next, we bring in the race of the victim by defining the *r.v.*

$$\underset{\sim}{V} = \begin{cases} \textit{white} & \text{victim is white} \\ \textit{black} & \text{victim is black} \end{cases}$$

- We have the following probabilities

$$P\left\{ \underset{\sim}{D} = 1 \,\middle|\, \underset{\sim}{R} = \textit{white}, \underset{\sim}{V} = \textit{white} \right\} = \frac{19}{151} = 0.126$$

$$P\left\{ \underset{\sim}{D} = 1 \,\middle|\, \underset{\sim}{R} = \textit{black}, \underset{\sim}{V} = \textit{white} \right\} = \frac{11}{63} = 0.175$$

DISCRIMINATION CASE STUDY: USING MORE DATA

$$P\left\{ \underset{\sim}{D} = 1 \,\middle|\, \underset{\sim}{R} = white, \underset{\sim}{V} = black \right\} = \frac{0}{9} = 0$$

$$P\left\{ \underset{\sim}{D} = 1 \,\middle|\, \underset{\sim}{R} = black, \underset{\sim}{V} = black \right\} = \frac{6}{103} = 0.058$$

- Data disaggregation on the basis of conditioning also on the *r.v.* $\underset{\sim}{V}$ shows that blacks appear to get the death penalty more frequently, about $5\,\%$ more, than whites independent of the race of the victim

APPARENT PARADOX

- **No difference** **between the overall imposition of death penalty and the race of the convicted murderers in the** **aggregated data case**
- **Clear difference** **in the** **disaggregated data case** **where the race of the victim is explicitly considered:** ***blacks*** **appear to get the penalty with 5*%* higher incidence than** ***whites***
- **The** **classification of the victim's race** **allows the distinct differentiation of the** $\underset{\sim}{R}$ = ***white*** **from the** $\underset{\sim}{R}$ = ***black*** **cases**

KEY ISSUE

- Since the number of *black* victims for $\underset{\sim}{R}$ = *white* cases is *0*, the result is a *0* rate of death penalty, making no contribution to the overall rate for the $\underset{\sim}{R} = white$ cases
- In addition, the many *black* victims for the $\underset{\sim}{R} = black$ cases results in the relatively low death penalty rate for *black* defendant / *black* victim cases and brings down the overall death penalty rate for *black* victims