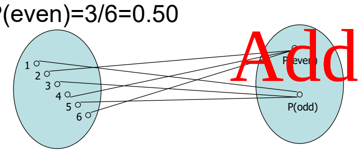
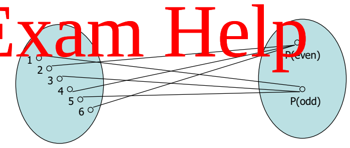
Probability and Statistics Review Part II

Decision Theory

- Events may be discrete (fixed number of outcomes) or continuous (infinite outcomes)
- Discrete events:
  - Number of truck arrivals to a receiving dock
  - Number of cases opened on “Deal or No Deal” out of 26 without opening the $1M case
  - Number of failures in a production lot of 1000 units
- Continuous events:
  - Mean time to failure of a component
  - Percent contamination level in a 100 cc sample of river water
  - Weight of quarter pound burger patties

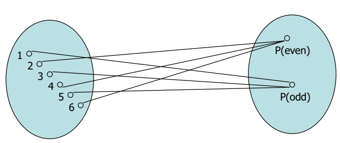
- Probabilities typically represented with functions that map events to a numerical probability
- Must still adhere to rules of probability
  - Ex: tossing 1 die (6 outcomes); event= even or odd
  - P(even)=3/6=0.50





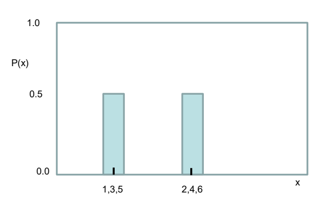
The function would be:

$$P(event) = \begin{cases} \frac{3}{6} & if \quad x = 2,4,6 \\ \frac{3}{6} & if \quad x = 1,3,5 \\ 0 & otherwise \end{cases}$$

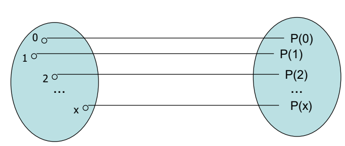


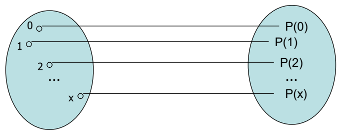
The graph would be:

$$P(event) = \begin{cases} \frac{3}{6} & if \quad x = 2,4,6 \\ \frac{3}{6} & if \quad x = 1,3,5 \\ 0 & otherwise \end{cases}$$



- Another example
- Number of arrivals to a shop Poisson distributed with average of 8 customers per hour

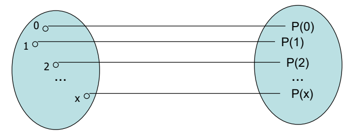




The function would be:

$$P(no.\ arrivals = x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \qquad x = 0,1,2\ldots,n$$

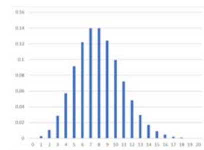
$= \frac{8^x e^{-8t}}{x!}$ (for example)



The graph format would be:

P(x)

This



x→

$P(no.\ arrivals = x) = \frac{8^x e^{-8t}}{x!}$

(for example)

Conditional probability:
"Event A depends on event B"

P(A|B) = joint probability of both A and B divided by the marginal probability of B or
= P(A and B)/P(B)

Is dependence the same as not mutually exclusive? (or independence the same as mutually exclusive?)

Statistical independence and dependence vs. mutually exclusive

- Given 2 events A and B, P(A|B)=P(A and B)/P(B)
- If mutually exclusive, P(A and B)=0 so P(A|B)=0 (A has nothing to do with B)
- Independence: when occurrence of one event has no bearing on the other event; P(A|B)=P(A) and P(A and B)=P(A)P(B)

Related note: if 2 events mutually exclusive they cannot be independent; ie if A, B mutually exclusive, if A occurs, B cannot occur so P(A|B)=0 $\neq$ P(A)

- Example: we choose a car at random; let A=4 cylinder engine and B=6 cylinder engine
- P(A) has some value > 0
- A and B are mutually exclusive
- But P(A|B)=0 $\neq$ P(A) (which is >0) so not independent. Saying car has 4 cylinders means it does not have 6.

Definitions:

Random variable. A variable used to represent the events associated with a sample space

Expected value. The mean (average), weighted value of a random variable based on its probability distribution

Variance. The weighted sum of differences of all points in the sample space from the mean. The standard deviation=the square root of the variance.

ISE 562; Dr. Smith

Calculating probabilities by relative frequency

$P(E) = n_E / N_S$

Ex: Compute the probability of one boy in a family of 3 children

N_s=8 outcomes in sample space

bbb bbg bgb <u>bgg</u>

gbb <u>gbg</u> <u>ggb</u> ggg

P(1 b)= 3/8

8/21/2022 13

ISE 562; Dr. Smith

Calculating probabilities-multiplication rule

Given k sets of n_k items, the number of possible cases = $n_1 \times n_2 \times \ldots n_k$

Ex: If we want to buy a computer with 3 available monitor types, m1, m2, m3; 4 CPU speeds, c1, c2, c3, c4; and 3 hard drive capacities, h1, h2, h3; what is the probability we randomly select the combination (m2, c3, h1)?

Number of outcomes = 3x4x3=36

Prob (m2, c3, h1)=1/36

8/21/2022 14

ISE 562; Dr. Smith

Calculating probabilities-multiplication rule

Could have used this on previous example to compute number in sample space:

Three sets each with 2 possibilities (b,g)

Child 1	Child 2	Child 3
b	b	b
g	g	g

Number of outcomes = 2x2x2=8

8/21/2022 15

ISE 562; Dr. Smith

Expected value and variance

Example: Suppose the random variable for the number of proposals funded can be represented by the pdf

Proposal	P(selection)
X=1	0.35
X=2	0.40
X=4	0.25

$$E(X) = \sum_{i=1}^{3} xP(X) = 1\cdot(0.35) + 2\cdot(0.40) + 4\cdot(0.25)$$
$$= 2.15$$
$$V(X) = \sum_{i=1}^{3} (x - E(X))^2 P(X) = (1-2.15)^2 \cdot (0.35)$$
$$+ (2-2.15)^2(0.40) + (4-2.15)^2(0.25) = 1.3275$$

8/21/2022 16

ISE 562; Dr. Smith

The binomial distribution—a discrete pdf

- Derived from the Bernoulli distribution
- Only 2 outcomes possible (success or failure)
- P(success) the same from trial to trial
- There are N trials (fixed)
- The N trials are independent
- Random variable is number of successes, r, in n trials

8/21/2022 17

ISE 562; Dr. Smith

The binomial distribution

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

p=probability of success; fixed over range of trials
n=number of independent trials
r=number of successes in n trials

8/21/2022 18

The binomial distribution

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

Example: In a family of three children, what is the probability of 1 boy? Assuming the probability of a boy = 0.50 we want to know P(r=1) with n=3 trials:

$P(1)=3!/(1!)(2!)\ (.5)^1\ (.5)^2 = 0.375$

(=3/8 as shown on slide 11)

The binomial distribution

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

Example: suppose the probability is 0.02 that a certain lab test will fail to detect a disease. What is the probability that among 20 such tests, 2 will fail?

$P(2)=20!/(2!)(18!)\ (.02)^2\ (.98)^{18} = 0.0528$

Also…

Example application

Biomorphic explorers for Mars exploration

Example

Bayes Theorem

Our hero!

Bayes Theorem

English theologian and mathematician Thomas Bayes has greatly contributed to the field of probability and statistics. His ideas have created much controversy and debate among statisticians over the years.
Thomas Bayes was born in 1702 in London, England. There appears to be no exact records of his birth date. Bayes's father was one of the first six Nonconformist ministers to be ordained in England. Bayes's parents had their son privately educated. There is no information about the tutors Bayes worked with. However, there has been speculation that he was taught by de Moivre, who was doing private tuition in London during this time. Bayes went on to be ordained, like his father, a Nonconformist minister. He first assisted his father in Holborn, England. In the late 1720's, Bayes took the position of minister at the Presbyterian Chapel in Tunbridge Wells, which is 35 miles southeast of London. Bayes continued his work as a minister up until 1752. He retired at this time, but continued to live in Tunbridge Wells until his death on April 17, 1761. His tomb is located in Bunhill Fields Cemetery in London. Throughout his life, Bayes was also very interested in he field of mathematics, more specifically, the area of probability and statistics. Bayes is believed to be the first to use probability inductively. He also established a mathematical basis for probability inference. Probability inference is the means of calculating, from the frequency with which an event has occurred in prior trials, the probability that this event will occur in the future. According to this Bayesian view, all quantities are one of two kinds: known and unknown to the person making he inference. Known quantities are obviously defined by their known values. Unknown quantities are described by a joint probability distribution. Bayesian inference is seen not as a branch of statistics, but instead as a new way of looking at the complete view of statistics. Bayes wrote a number of papers that discussed his work. However, the only ones known to have been published while he was still living are: *Divine Providence and Government Is the Happiness of His Creatures* (1731) and *An Introduction to the Doctrine of Fluxions, and a Defense of the Analyst* (1736). The latter paper is an attack on Bishop Berkeley for his attack on the logical foundations of Newton's Calculus. Even though Bayes was not highly recognized for his mathematical work during his life, he was elected a Fellow of the Royal Society in 1742.

Bayes Theorem

Perhaps Bayes's most well known paper is his *Essay Towards Solving a Problem in the Doctrine of Chances.* This paper was published in the *Philosophical Transactions of the Royal Society of London* in 1764. This paper described Bayes's statistical technique known as Bayesian estimation. This technique based the probability of an event that has to happen in a given circumstance on a prior estimate of its probability under these circumstances. This paper was sent to the Royal Society by Bayes's friend Richard Price. Price had found it among Bayes's papers after he died. Bayes's findings were accepted by Laplace in a 1781 memoir. They were later rediscovered by Condorcet, and remained unchallenged. Debate did not arise until Boole discovered Bayes's work. In his composition the *Laws of Thought*, Boole questioned the Bayesian techniques.
Boole's questions began a controversy over Bayes's conclusions that still continues today. In the 19th century, Laplace, Gauss, and others took a great deal of interest in this debate. However, in the early 20th century, this work was ignored or opposed by most statisticians. Outside the area of statistics, Bayes continued to have support from certain prominent figures. Both Harold Jeffreys, a physicist, and Arthur Bowley, an econometrician, continued to argue on behalf of Bayesian ideas. The efforts of these men received help from the field of statistics beginning around 1950. Many statistical researchers, such as L. J. Savage, Buno do Finetti, Dennis Lindley, and Jack Kiefer, began advocating Bayesian methods as a solution for specific deficiencies in the standard system.

However, some researchers still argue that concentrating on inference for model parameters is misguided and uses unobservable, theoretical quantities. Due to this skepticism, some are reluctant to fully support the Bayesian approach and philosophy.

ISE 562; Dr. Smith

Bayes Theorem

Bayes is buried in Bunhill Fields in the heart of the City of London. The cemetery was used for the burial of nonconformists in the 18th century, but is now a public park maintained by the Corporation of London. Also buried in Bunhill Fields is Bayes's friend Richard Price, a pioneer of insurance, who presented Bayes's famous paper on probability to the Royal Society in 1763, two years after Bayes's death. Across the City Road from Bunhill Fields is Wesley's Chapel, which has been restored in recent years. The pictures below show Bayes's tomb with a variety of inscriptions. It was a family vault in which are laid several members of the Bayes, Cotton and West families. On the top of the tomb is an inscription saying how the tomb was restored in 1969, through public subscription from statisticians worldwide." These photos were taken by Professor Tony O'Hagan of Sheffield University who also also provided the information about the burial place. Bunhill (probably a corruption of "bonehill") Fields operated as a burial ground for "Dissenters" from 1665 to 1853, during which time around 123,000 burials took place. There are many notable graves, including John Bunyan, William Blake, Daniel Defoe, many of the Cromwell family and Susanna Wesley (mother of John Wesley, the founder of Methodism, who is buried across the City Road where his chapel still stands).

8/21/2022 25

ISE 562; Dr. Smith

Bayes Theorem—derivation

First some notation:

- P(events A and B occurring) = $P(A \cap B) = P(A,B)$
- P(A and not B) = P(A,not_B)
- P(A given B) = P(A|B)

8/21/2022 26

ISE 562; Dr. Smith

Bayes Theorem—derivation

Start with 2 probability rules:

i) P(Y|X) = P(Y,X)/P(X) = P(X,Y)/P(X)

ii) P(Y)=P(Y,X)+P(Y,not_X)=P(X,Y)+P(not_X,Y)

- In a decision problem we are given X and P(X); Y and P(Y|X);
- Desire P(X|Y) (posterior) as a function of what is known
- Substituting (ii) into (i) we get:

iii) P(X|Y)=P(X,Y)/[P(X,Y)+P(not_X ,Y)]; if we don't know the denominator terms we can calculate them from (i):

iv) P(X,Y)=P(Y|X) P(X); and P(not_X,Y) = P(Y|not_X) P(not_X)

Substituting (iv) back into denominator of (iii) yielding Bayes Rule

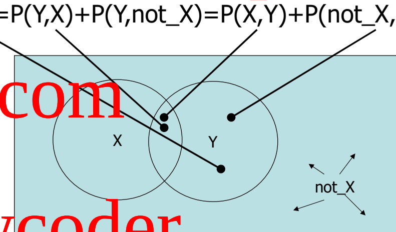
P(X|Y)=P(Y|X) P(X)/ [P(Y|X)P(X)+P(Y|not_X)P(not_X)]

8/21/2022 27

ISE 562; Dr. Smith

Graphic Proof of rule ii

i) P(Y)=P(Y,X)+P(Y,not_X)=P(X,Y)+P(not_X,Y)



8/21/2022 28

ISE 562; Dr. Smith

To get the discrete form, we add an index, for the state of X:

Prior distribution (original state of knowledge about Xr)

Likelihood distribution from new information (sample)

$$P(X_r \mid Y) = \frac{P(X_r)P(Y \mid X_r)}{\underbrace{\sum_i P(X_i)P(Y \mid X_i)}_{\text{Normalizing constant}}}$$

8/21/2022 29

ISE 562; Dr. Smith

Bayes Theorem—example

Three facilities supply plastic containers to a manufacturer. All are made to same specification. However, after months of testing, records indicate the following:

Prior(X)

Supplying facility	Fraction supplied by	Fraction defective
1	0.15	0.02
2	0.80	0.01
3	0.05	0.03

8/21/2022 30

Bayes Theorem—example

The director of manufacturing randomly selects a unit, has it tested, and finds it to be defective. Let Y=event that item is defective and X_i be the event the item came from facility i=1, 2, 3. Use Bayes rule to determine the probability the defective came from facility 1, 2, or 3 given it was defective. That is, $P(X_1|Y)$, $P(X_2|Y)$, and $P(X_3|Y)$.

$$P(X_r|Y) = \frac{P(Y|X_r)P(X_r)}{\sum_r P(Y|X_i)P(X_i)}$$

Bayes Theorem—example

$$P(X_1|Y) = \frac{P(Y|X_1)P(X_1)}{P(X_1)P(Y|X_1)+P(X_2)P(Y|X_2)+P(X_3)P(Y|X_3)}$$

$$= \frac{(0.02)(0.15)}{(0.02)(0.15)+(0.01)(0.80)+(0.03)(0.05)} = 0.24$$

$$P(X_2|Y) = \frac{(0.01)(0.80)}{(0.02)(0.15)+(0.01)(0.80)+(0.03)(0.05)} = 0.64$$

$$P(X_3|Y) = \frac{(0.03)(0.05)}{(0.02)(0.15)+(0.01)(0.80)+(0.03)(0.05)} = 0.12$$

	$P(X_r)$	$P(Y\|X_r)$	$P(X_r\|Y)$
Supplier, X_r	% supplied by company X_r	Fraction defective	New estimate of P(1,2, or 3 the source of defective)
1	0.15	0.02	0.24
2	0.80	0.01	0.64
3	0.05	0.03	0.12

Initially, **company 3 looked** the worst (.03 or 3 times worse than company 2). After sampling, company 3 is the best and company 2 is the worst (>5 times worse than company 3!) Why? Because company 2 accounts for 80% of the product volume—information not available in $P(X_r)$

Another Example

- The following categories are from an online retail shopping site you are opening. You want to track the customer's interest in each category so you can advertise similar items when they make future visits.
- You make the assumption that the number of searches = the level of interest in a category so you can estimate the probability the customer will be interested in that category in the future.
- Let x= buys item in category x and let y = random variable representing "interested in category y"
- Let x and y index over Arts&Crafts, Grocery, Sports, Entertainment

- Next we need the Prob(buy x) but we have no data so assume a diffuse prior (uniform pdf)

Category, x	P(buy x)
Arts & Crafts	0.2500
Grocery	0.2500
Sports	0.2500
Entertainment	0.2500

- We need the likelihood of interest in x given a buy x, P(interest in x | buy x) but don't have any buy data yet so assume independence (P(interest in x|buy x)= P(interest in x). We use the counts of searches in each category:

Category, x	# searches	P(interest in x)
Arts & Crafts	4	0.2500
Grocery	6	0.3750
Sports	5	0.3125
Entertainment	1	0.0625
	16	

ISE 562; Dr. Smith

- Now compute the posterior pdf of P(buy x | interest in y) using Bayes theorem

$$P(X_r|Y) = \frac{P(Y|X_r)P(X_r)}{\sum_r P(Y|X_i)P(X_i)}$$

Category	Prior(buy in x)	Likelihood(interest in y|buy in		Like x Prior	Posterior(buy x | interest in y)
Arts & Crafts	0.2500	0.250		0.0625	0.2500
Grocery	0.2500	0.375		0.0938	0.3750
Sports	0.2500	0.313		0.0781	0.3125
Entertainment	0.2500	0.063		0.0156	0.0625
			Sum =	0.25	1

- Note the posterior is the <u>same</u> as the likelihoods (the only information available—but there was <u>no</u> information in the uniform diffuse prior!).

8/21/2022 37

ISE 562; Dr. Smith

- Now the customer buys 1 item from Arts & Crafts and 2 items from Grocery and 1 from Entertainment. This is new information to update the prior.
- We have data on buys so compute Likelihood of interest given buys in x as #buys in category x ÷ #searches in category x. (note: if # buys or # searches = zero, likelihood = 0).

Bought in Category, y	# searches	# buys in x	Likelihood(interest in y|buy in x)
Arts & Crafts	4	1	0.250
Grocery	6	2	0.333
Sports	5	0	0.000
Entertainment	1	1	1.000
	16		

8/21/2022 38

ISE 562; Dr. Smith

- Using the likelihoods and prior distributions above, compute the posterior probability for interest in x given purchases y so items in x can be prioritized for presentation on the web pages.

The 2 buys in Grocery drove the P(buy | interest) probability from 0.375 to 0.50

Category	Prior(buy in x)	Likelihood(interest|buy in x)	Like x Prior	Posteri[illegible] x|interest in in x)
Arts & Crafts	0.2500	0.250	0.0625	0.2500
Grocery	0.3750	0.333	0.1250	0.5000
Sports	0.3125	0.000	0.0000	0.0000
Entertainment	0.0625	1.000	0.0625	0.2500

8/21/2022 39

ISE 562; Dr. Smith

- Now suppose the user returns and conducts 1 new search with another purchase in Entertainment.

Bought in Category, y	# searches	# buys	Likelihood(buy y | interest in x)
Arts & Crafts	4	1	0.250
Grocery	6	2	0.333
Sports	5	0	0.000
Entertainment	2	2	1.000
	17		

- Now update the posterior with these new data:

8/21/2022 40

ISE 562; Dr. Smith

- Now update the posterior with these new data:

# searches	# buys
4	1
6	2
5	0
2	2
17	

The second buy in Entertainment drove the P(buy | interest) probability from 0.25 to 0.52

Category	Prior(buy in x)	Likelihood(interest|buy in x)		Like x Prior	[illegible]or(interest in x | buy
Arts & Crafts	0.2500	0.250		0.0625	0.1304
Grocery	0.5000	0.333		0.1667	0.3478
Sports	0.0000	0.000		0.0000	0.0000
Entertainment	0.2500	1.000		0.2500	0.5217
			Sum =	0.4792	

- The focus moves away from Grocery category to Entertainment with same number of buys but with more intense interest (2 buys in 2 searches vs. Grocery with 2 buys and 6 searches implying searches in Entertainment more productive).

8/21/2022 41

ISE 562; Dr. Smith

The normal distribution—a continuous pdf

- Representative of many natural processes
- Standardized in tabular form
- Probability defined for intervals, not points
- Requires mean and variance to calculate probability
- Lookup values found using $z=(x-\mu)/\sigma$
  - μ=mean; σ=standard deviation

8/21/2022 42

The normal distribution—a continuous pdf

- Example: The cost per patient for a particular medical procedure was determined from records to be normally distributed with mean $200 and standard deviation +/- $50.
- For a random selection of records, what is the probability the cost is less than $150?
- What is the probability the cost is between $180 and $220 per procedure?

The normal distribution—a continuous pdf

- $P(X<150)=P_N(z< (150\text{-}200)/50)$
  $=P(z< -1)=0.16$ (from table)
- $P(180<X<220)$
  $=P(z<(220\text{-}200)/50) - P(z<(180\text{-}200)/50$
  $=P(z<0.40) – P(z<-0.40)$
  $=0.6554 – 0.3446$
  $=0.3108$

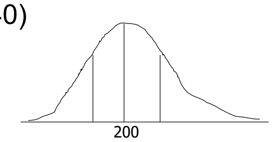


Table TI-83

Discrete Bayes Methods

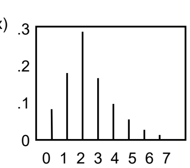
Value and Decision Theory

A little more review of discrete probability…

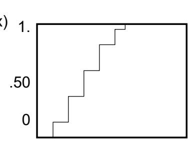
Discrete distributions

- Let X= discrete random variable of the number of rainy days in Chicago next week

	PMF	CDF
X	P(X=x)	P(X≤x)
0	.13	.13
1	.19	.32
2	.29	.61
3	.18	.79
4	.11	.90
5	.06	.96
6	.03	.99
7	.01	1.00



PMF



CDF

Discrete distributions

Expected value of distribution

$$E[X]=\sum_{i=0}^{7} X_i P(X_i)=2.3\ days$$

Expected value of a function: suppose the profit to an amusement park is -2 if it rains and +12 if no rain.

$$Cost(X)=\begin{cases} 12\ if\ \ X=0 \\ -2X\ \ if\ \ X>0 \end{cases}$$

$$E[Cost(X)]=\sum_{i=0}^{7} Cost(X_i)P(X_i)$$
$$=12(.13)-2(.19)-4(.29)-6(.18)-8(.11)-10(.06)$$
$$-12(.03)-14(.01)=-2.28$$

Laws of expectations

- E(c)= c
- E(cX)=cE(x); E(c+X)=c+E(X)
- E(X+Y)=E(X)+E(Y)
- E(aX+bY)=aE(X)+bE(Y)
- $E(\Sigma c_i X_i) = \Sigma c_i E(X_i)$

- Variance

$$V[X]=\sigma^2=E[(X-E[X])^2]=\sum_{i=0}^{7}(X_i-E[X])^2P(X_i)$$

$$=\sum_{i=0}^{7}(X_i-2.3)^2P(X_i)=2.51$$

$\sigma=\pm\sqrt{\sigma^2}$ (Standard deviation)

Laws of variance

- $V[X]=E[X^2]-[E[X]]^2$
- $V[cX]=c^2V[X]$
- V[c]=0
- V[X+c]=V[X]

- Joint probability distributions (more than 1 random variable), $P(X_1,X_2,\ldots,X_n)$
- For 2 random variables, $P(X_1,X_2)$
- Graph is 3-dimensional: X_1 vs X_2 vs $P(X_1,X_2)$
- Usually displayed in table form

Joint Probability Table

$P(S_i)$

	C1=Science	C2=Mgmt	C3=Engr.	Total
S1=Male	.250	.350	.200	.800
S2=Female	.100	.050	.050	.200
Total	.350	.400	.250	1.000

$P(C_j)$

Joint Probabilities

Marginal Probabilities

Back to the iterative nature of Bayes rule

- Again, we are concerned about fraction defectives for a mfg process.
- We represent the fraction defectives with variable, p.
- 4 states of nature believed possible:
  1. no malfunctions (p=.01)
  2. type x malfunction (p=.05)
  3. both a type x and y malfunction (p=.10)
  4. a type x, y, and z malfunction (p=.25)

Example

- The manager subjectively estimates the probabilities for these fraction defectives as:

  P(p=.01)=.60
  P(p=.05)=.30
  P(p=.10)=.08
  P(p=.25)=.02

  This is the manager's *prior* distribution for the parameter, p.

- The manager also wants some solid data observations by sampling 5 units from the process.
- One defective is found
- The repetitive nature of the process indicates independence between units, constant probabilities of success (Bernoulli trials), and two outcomes (success or failure)
- Let "success" = defective unit
- We can use binomial pdf to calculate the likelihoods of observing the sample result

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

P(r=1|n=5,p=.01)=5!/(1! (4!)) $(.01)^1(.99)^4$=.0480
P(r=1|n=5,p=.05)=5!/(1! (4!)) $(.05)^1(.95)^4$=.2036
P(r=1|n=5,p=.10)=5!/(1! (4!)) $(.10)^1(.90)^4$=.3280
P(r=1|n=5,p=.25)=5!/(1! (4!)) $(.25)^1(.75)^4$=.3955

These are the likelihoods of observing the sample of size 5 with one defective. Let's put in table form to view the calculation of the posterior pdf by Bayes rule

$$posterior\ probability = \frac{(prior\ prob.)(likelihood)}{\sum (prior\ prob.)(likelihood)}$$

p (Possible states of nature)	Prior prob. (Given)	Likelihood (Computed)	Prior prob * likelihood (Numerator (col 2 x col 3))	Posterior prob. (Revised pdf)
.01	.60	.0480	.02880	.232
.05	.30	.2036	.06108	.492
.10	.08	.3280	.02624	.212
.25	.02	.3955	.00791	.064
	1.00		.12403	1.000

Most likely

denominator

Now the manager decides to take another sample—we use the posterior pdf as the updated prior to reflect the first sample

  P(p=.01)=0.232
  P(p=.05)=0.492
  P(p=.10)=0.212
  P(p=.25)=0.064

  This is the manager's new *prior* distribution for the parameter, p.

- The manager wants more observations and samples another 5 units from the process.
- Two defectives are found
- We again use the binomial pdf to calculate the likelihoods of observing the sample result

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

P(r=2|n=5,p=.01)=5!/(2! (3!)) $(.01)^2(.99)^3$=.0010
P(r=2|n=5,p=.05)=5!/(2! (3!)) $(.05)^2(.95)^3$=.0214
P(r=2|n=5,p=.10)=5!/(2! (3!)) $(.10)^2(.90)^3$=.0729
P(r=2|n=5,p=.25)=5!/(2! (3!)) $(.25)^2(.75)^3$=.2637

These are the likelihoods of observing the sample of size 5 with two defectives. Let's put in table form to view the calculation of the posterior pdf by Bayes rule

After 2nd sample:

$$posterior\ probability = \frac{(prior\ prob.)(likelihood)}{\sum (prior\ prob.)(likelihood)}$$

p	Prior prob.	Likelihood	Prior prob * likelihood	Posterior prob.
.01	.232	.0010	.00023	.005
.05	.492	.0214	.01053	.244
.10	.212	.0729	.01545	.359
.25	.064	.2637	.01688	.392
	1.00		.04309	1.000

Most likely

- What if we had taken one sample of size 10 and observed the 3 defects?

We start with the original prior estimates

- The manager subjectively estimates the probabilities for these fraction defectives as:
  P(p=.01)=.60
  P(p=.05)=.30
  P(p=.10)=.08
  P(p=.25)=.02

  This is the manager's *prior* distribution for the parameter p

- A sample of size 10 is taken
- Three defectives found
- We can use the binomial pdf to calculate the likelihoods of observing the sample result

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

P(r=3|n=10,p=.01)=10!/(3! (7!)) $(.01)^3(.99)^7$=.0001
P(r=3|n=10,p=.05)=10!/(3! (7!)) $(.05)^3(.95)^7$=.0105
P(r=3|n=10,p=.10)=10!/(3! (7!)) $(.10)^3(.90)^7$=.0574
P(r=3|n=10,p=.25)=10!/(3! (7!)) $(.25)^3(.75)^7$=.2503

These are the likelihoods of observing the sample of size 10 with 3 defectives. Let's put in table form to view the calculation of the posterior pdf by Bayes rule

$$posterior\ \ probability = \frac{(prior\ prob.)(likelihood)}{\sum (prior\ prob.)(likelihood)}$$

p	Prior prob.	Likelihood	Prior * likelihood	Posterior prob.	(2 sample case)
.01	.60	.0001	.00006	.005	.005
.05	.30	.0105	.00315	.245	.244
.10	.08	.0574	.00459	.359	.359
.25	.02	.2503	.00501	.391 Most likely	.392
	1.00		.01281	1.000	1.000

The answer is the same!

(rounding error in 3rd decimal)

- Poisson Example

$$P(r\,|\,t,\lambda) = \frac{e^{-\lambda t}(\lambda t)^r}{r!} \qquad r = 1,2,\ldots$$

- Events are numbers of occurrences over a fixed continuum (usually time but can be counts, area, volume, network server hits, etc.)
- Time interval analogous to a trial
- Lambda represents the average number of occurrences in time interval

- Poisson Example
- Car sales rep performance characterized as poor, good, great by selling cars at a rate of one every 8th, 4th, and 2 days, respectively. The sales manager wants to hire a new rep.
- Let λ be the random variable to represent the average sales rate per day. Converting the above values to a "per day" basis, the three possible outcomes of the new hire in terms of λ become:

λ= 1/8, ¼, and ½ .

- Poisson Example

The sales manager estimates the prior distribution for λ as:

$P(\lambda=1/8)=0.30$, $P(\lambda=1/4)=0.50$, $P(\lambda=1/2)=0.20$

A sample is observed and there are 10 sales in 24 days so r=10 and t=24

We have to calculate the Poisson probabilities for each λ value

- Poisson Example

$$P(r=10\,|\,t=24,\lambda=\frac{1}{8}) = \frac{e^{-3}(3)^{10}}{10!} = .0008$$

$$P(r=10\,|\,t=24,\lambda=\frac{1}{4}) = \frac{e^{-6}(6)^{10}}{10!} = .0413$$

$$P(r=10\,|\,t=24,\lambda=\frac{1}{2}) = \frac{e^{-12}(12)^{10}}{10!} = .1048$$

$$posterior\ \ probability = \frac{(prior\ prob.)(likelihood)}{\sum (prior\ prob.)(likelihood)}$$

$$P(\lambda_j\,|\,y) = \frac{P(\lambda_j)P(y\,|\,\lambda_j)}{\sum_i P(\lambda_i)P(y\,|\,\lambda_i)}$$

Before sampling, the odds of new employee being a great salesman were 1 in 5 (20%). After selling 10 cars in 24 days, the odds are even (50-50) reflecting better than expected performance (manager's prior) of the employee.

λ	Prior prob.	Likelihood	Prior prob * likelihood	Posterior prob.
1/8	.30	.0008	.00024	.006
1/4	.50	.0413	.02065	.493
1/2	.20	.1048	.02096	.501 Most likely
	1.00		.04185	1.000

- If another sample was taken, the posterior pdf would be used as the prior for the next iteration.
- The procedure is the same for other discrete probability distributions
- You can use the table format or the formula format—your preference

- *What if both distributions are tabular?*

- You ask the software development team for an estimate to complete a task.
  - They estimate an 80% chance it will be done in <3 months (20% > 3 months).
- From past history of other teams' (data = schedule records) the likelihood of completion as predicted was:
  - 40% that it was on time and 60% that the delivery exceeded the earlier estimate
  - Consider if we had 1, 2, or 3 sample schedules (i.e., 1, 2, or 3 pieces of evidence)

- The likelihood probability for N = 1 sample
  - Likelihood(t<3months) = 0.40
  - Likelihood(t>3months) = 0.60
- The likelihood probability for N = 2 samples
  - Likelihood(t<3months) = $(0.40)^2$ = 0.16
  - Likelihood(t>3months) = $(0.60)^2$ = 0.36
  - The likelihood probability for N = 3 samples
  - Likelihood(t<3months) = $(0.40)^3$ = 0.064
  - Likelihood(t>3months) = $(0.60)^3$ = 0.216

- For sample of N=1

θ	Prior prob.	Likelihood	Prior prob x likelihood	Posterior prob.
<3	0.8	0.40	0.32	0.727
>3	0.2	0.60	0.12	0.273
	1.00		0.44	1.000

- For sample of N=2

θ	Prior prob.	Likelihood	Prior prob x likelihood	Posterior prob.
<3	0.8	0.16	0.128	0.640
>3	0.2	0.36	0.072	0.360
	1.00		0.200	1.000

- For sample of N=3

θ	Prior prob.	Likelihood	Prior prob x likelihood	Posterior prob.
<3	0.8	0.064	0.0512	0.542
>3	0.2	0.216	0.0432	0.458
	1.00		0.0944	1.000

- Result is the *same* as if we took a sample of 1 three times:

θ	Prior prob.	Likelihood	Prior prob x likelihood	Posterior prob.
<3	0.8	0.40	0.32	0.727
>3	0.2	0.60	0.12	0.273
	1.00		0.44	1.000

These stay the same as sample of N=1

Then use the posterior as the prior in 2nd sample

- Result is the same as if we took a sample of 1 three times:

θ	Prior prob.	Likelihood	Prior prob x likelihood	Posterior prob.
<3	0.727	0.40	0.291	0.640
>3	0.273	0.60	0.164	0.360
	1.00		0.455	1.000

These stay the same as sample of N=1

Then use the posterior as the prior in 3rd sample

- Result is the same as if we took a sample of 1 three times:

θ	Prior prob.	Likelihood	Prior prob x likelihood	Posterior prob.
<3	0.64	0.40	0.256	0.542
>3	0.36	0.60	0.216	0.458
	1.00		0.472	1.000

These stay the same as sample of N=1

Same result as sample of size N=2

- If another sample was taken, the posterior pdf would be used as the prior for the next iteration.
- Note that if the likelihoods are multiplied by any constants the posterior is unaffected because the same constants appear in the denominator.
- Suppose we double the likelihoods for N=2 (multiply by 2)…

- We get:

θ	Prior prob.	Likelihood	Prior prob x likelihood	Posterior prob.
<3	0.8	0.16x 2= 0.32	0.256	0.64
>3	0.2	0.36x 2= 0.72	0.144	0.36
	1.00		0.400	1.000

Same result as N=2

- Only the *ratio* between likelihoods affects the posterior

- Recap
- Bayes theorem provides a quantitative way to update the probability distribution of a decision variable as new data become available.
- If additional data are collected, the computed posterior distribution is used as the new prior in an iterative fashion.
- The posterior distribution probabilities will update after each round of new information.