Chapter 5:
Probability

Halima Bensmail

CS502

Monday 4-7pm

LAS Hall C

Random events frequently behave in ways that require complicated distributions. However, there are relatively simple distributions that are useful in many situations. One example is the **binomial distribution**.

The following are situations in which the binomial distribution might be used:

- How many heads in **10** tosses on a coin?
- How many democrats in a sample of **100 UT** students?

We will use the Binomial distribution in more complicated contexts, even if we cannot verify that these assumptions hold. E.g.

- Percent of defectives in a day's production

Bernoulli Distribution

Example: Toss of coin

Define $X = 1$ if head comes up and

$X = 0$ if tail comes up.

Both realizations are equally likely: $\mathbb{P}(X = 1) = \mathbb{P}(X = 0) = \frac{1}{2}$

Examples:

Often: Two outcomes which are *not* equally likely:

- Success of medical treatment
- Interviewed person is female
- Student passes exam
- Transmittance of a disease

Bernoulli distribution (with parameter θ)

- X takes two values, 0 and 1, with probabilities p and $1-p$
- Frequency function of X

$$p(x) = \begin{cases} \theta^x(1-\theta)^{1-x} & \text{for } x \in \{0,1\} \\ 0 & \text{otherwise} \end{cases}$$

- Often:

$$X = \begin{cases} 1 & \text{if event } A \text{ has occured} \\ 0 & \text{otherwise} \end{cases}$$

It means Bernouilli has two modes: p and 1-p

Example: A = blood pressure above 140 / 90 mm HG

Binomial distribution

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables

- Often only interested in number of successes

$$Y = X_1 + \ldots + X_n$$

We know:

- X_i is Bernoulli distributed with parameter θ
- X_i's are independent

What is the distribution of Y?

- Probability of realization $x_1, \ldots, x_n$ with y successes:

$$p(x_1, \ldots, x_n) = \theta^y (1 - \theta)^{n-y}$$

- Number of different realizations with y successes: $\binom{n}{y}$

THE BINOMIAL SETTING

1. There are a fixed number n of observations.
2. The n observations are all **independent.** That is, knowing the result of one observation tells you nothing about the other observations.
3. Each observation falls into one of just two categories, which for convenience we call "success" and "failure."
4. The probability of a success, call it p, is the same for each observation.

BINOMIAL DISTRIBUTION

The distribution of the count X of successes in the Binomial setting is the **Binomial distribution** with parameters n and p. The parameter n is the number of observations, and p is the probability of a success on any one observation. The possible values of X are the whole numbers from 0 to n.

The following is the formula used to compute probabilities for a binomial random variable

BINOMIAL PROBABILITY

If X has the Binomial distribution with n observations and probability p of success on each observation, the possible values of X are 0, 1, 2, . . . , n. If k is any one of these values,

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Computer Software and Tables are easier and faster ways to compute Binomial probabilities.

n! or (read as: factorial n)

$\binom{n}{k}$: is read as: n choose k

- $\binom{n}{k}$ =n!/k!(n-k)!
- Example:
- $\binom{5}{2}$=5!/2!(5-2)!
- 5!=…..?

- $P_{(k \text{ out of } n)} (X=k)=[n!/k!(n-k)!](p^k)(q^{n-k})$

where:

- n = the number of opportunities for event x to occur.
- k = the number of times that event x occurs.
- p = the probability that event x will occur on any particular occasion.
- q = the probability that event x will not occur on any particular occasion.

- In tossing a coin 5 times, what is the probability of ending up with exactly 2 heads among the 5 tosses?????

- n = 5 [the number of opportunities for a head to occur]
- k = 2 [the stipulated number of heads]
- p = .5 [the probability that a head will occur on any particular toss]
- q = .5 [the probability that a head will not occur on any particular toss]

Calculation by hand

- $P_{(k\ out\ of\ n)}$ = $[n!/k!(n-k)!](p^k)(q^{n-k})$
- P(2 out 5)=5!/[2!(5-2)!] $(0.5^2)(0.5^{5-2})$ =120/2x6 x (0.03125) =0.3125

R Documentation

The Binomial Distribution

function, quantile function and random generation for the binomial distribution with parameters size and prob.

```
rob, log = FALSE)
rob, lower.tail = TRUE, log.p = FALSE)
rob, lower.tail = TRUE, log.p = FALSE)
rob)
```

of quantiles.

of probabilities.

of observations. If length(n) > 1, the length is taken to be the number required.

of trials (zero or more).

ility of success on each trial.

if TRUE, probabilities p are given as log(p).

if TRUE (default), probabilities are $P[X \le x]$, otherwise, $P[X > x]$.

tion with size = n and prob = p has density

$$p(x) = choose(n,x) p^x (1-p)^{(n-x)}$$

that binomial *coefficients* can be computed by choose in R.

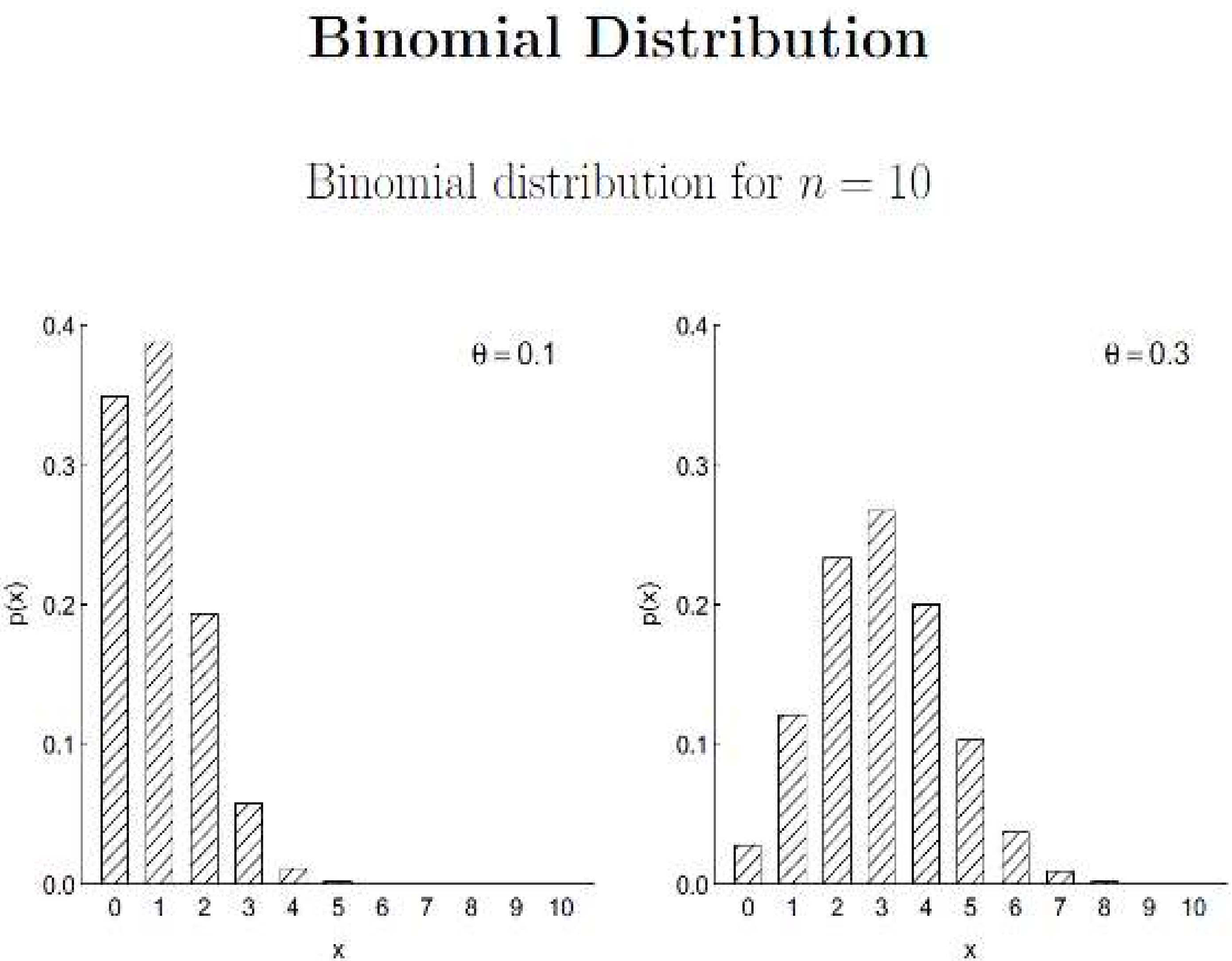
not integer, the result of dbinom is zero, with a warning. $p(x)$ is computed using Loader's algorithm, see the reference below.

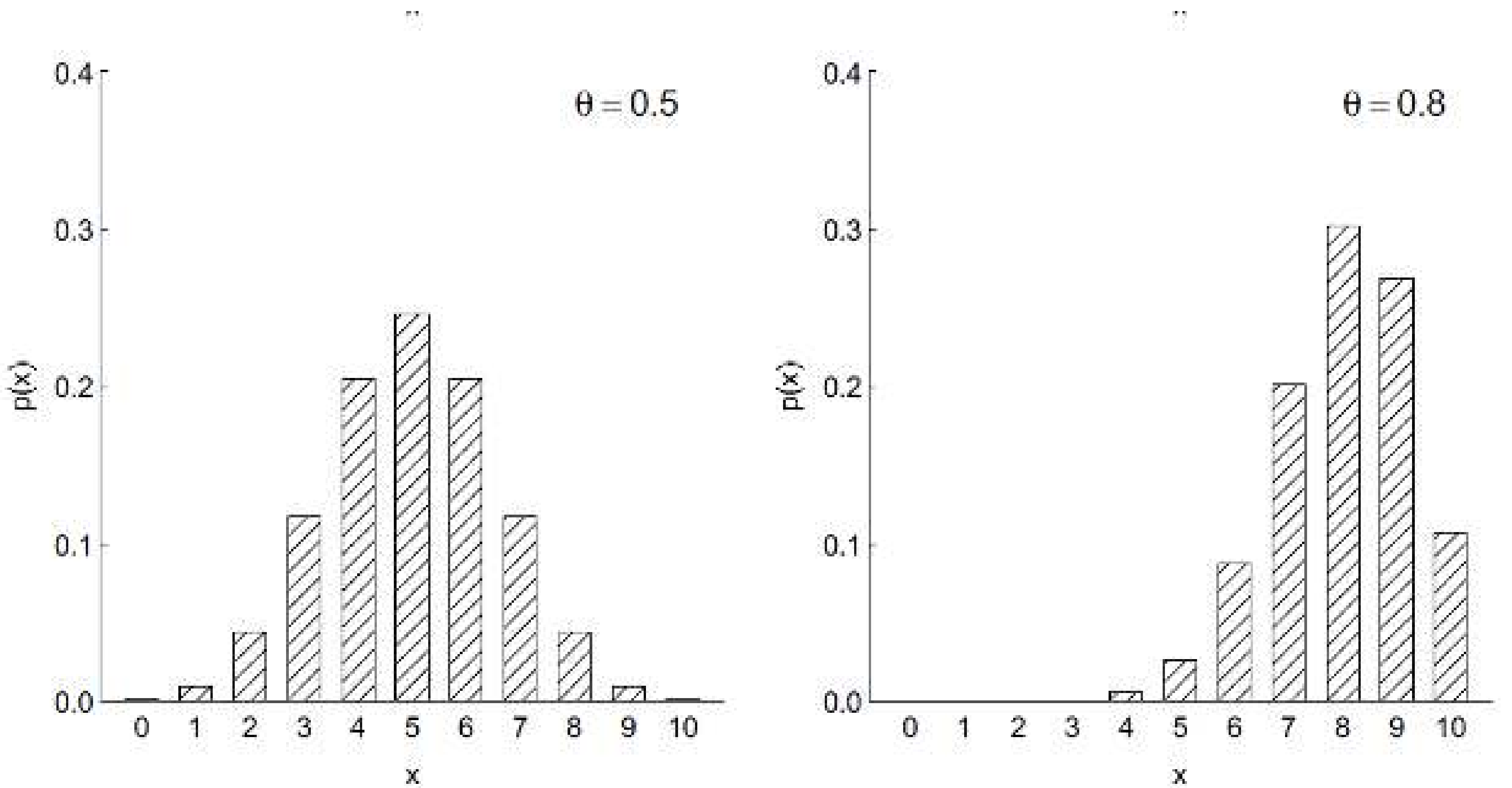
ed as the smallest value x such that $F(x) \ge p$, where F is the distribution function.

sity, pbinom gives the distribution function, qbinom gives the quantile function and rbinom generates random deviates.

ger, NaN is returned.

The distribution of the binomial Random variable with n = 10, p = 0.1





θ = 0.5
p(x)
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10
x
θ = 0.8
p(x)
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10
x

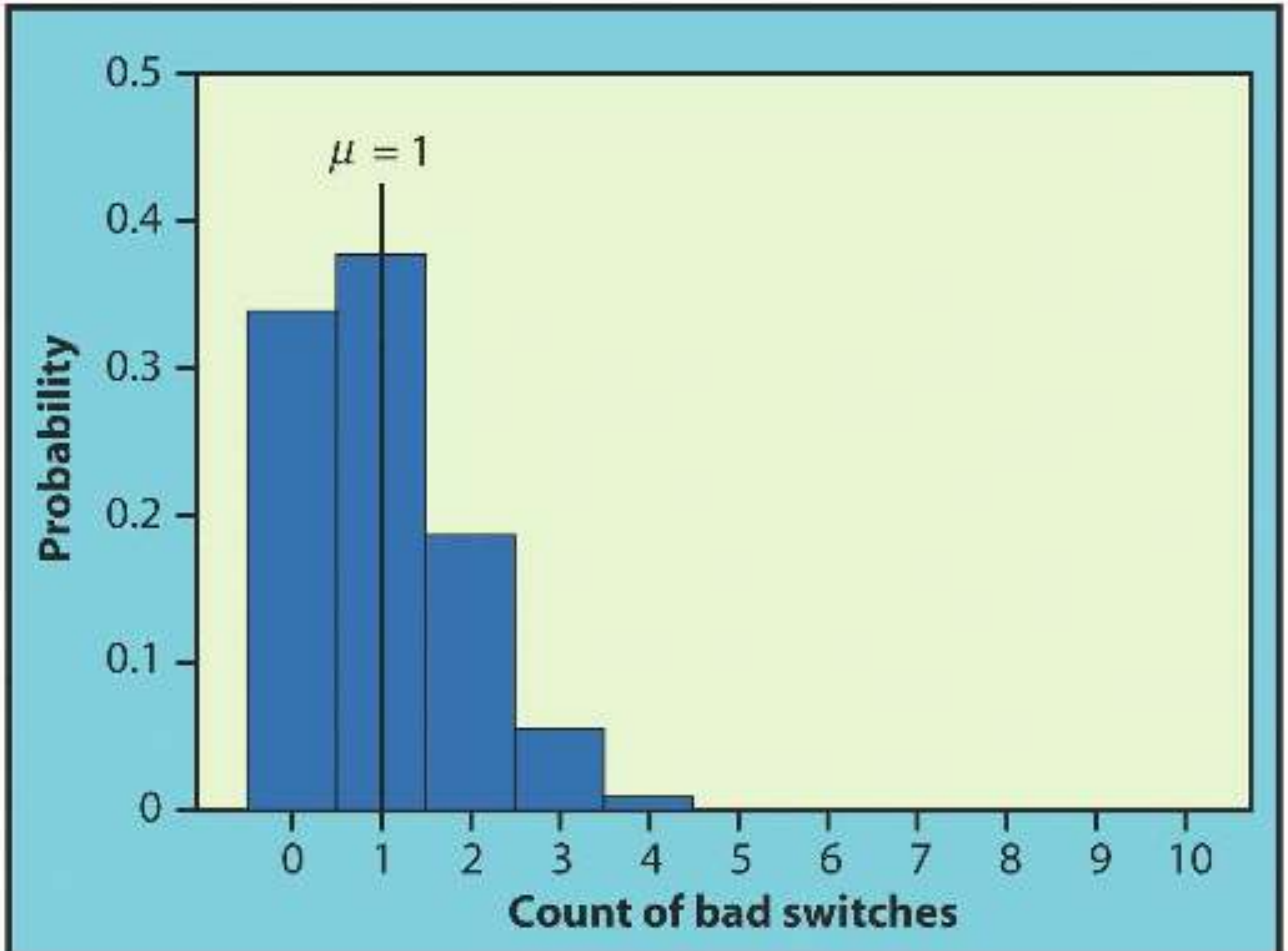
The mean and the standard deviation of a binomial are simple functions of n and p

BINOMIAL MEAN AND STANDARD DEVIATION

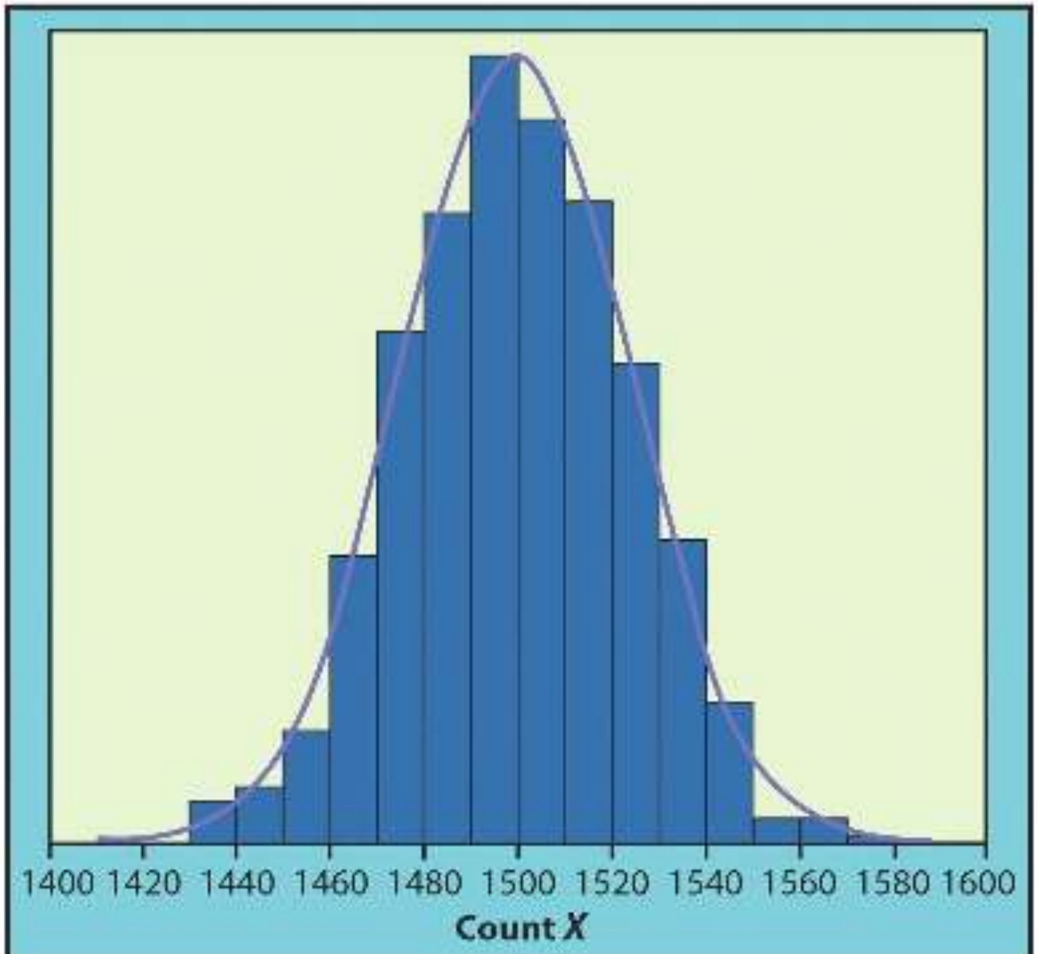
If a count X has the Binomial distribution with number of observations n and probability of success p, the **mean** and **standard deviation** of X are

$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

Find the mean when n = 10 and p = 0.1



Distribution of a binomial when n is large (= 2,500)
Any resemblance?



Using the **central limit theorem** it can be shown that, when **n** is large the binomial tends to behave like a normal distribution. This result is useful because for **large n** the calculations required to find probabilities for the binomial become computationally difficult. In contrast, finding probabilities for any normal distribution only involve simple tables.

NORMAL APPROXIMATION FOR BINOMIAL DISTRIBUTIONS

Suppose that a count X has the Binomial distribution with n trials and success probability p. When n is large, the distribution of X is approximately Normal, $N(np, \sqrt{np(1-p)})$.

As a rule of thumb, we will use the Normal approximation when n and p satisfy $np \geq 10$ and $n(1-p) \geq 10$.

Approximation of the Binomial: Poisson

θ is small
n is large $\Rightarrow$ *Binomial* becomes *Poisson* with rate λ

In fact:

$\lambda = n\theta$

GENERAL MULTIPLICATION RULE FOR ANY TWO EVENTS

The probability that both of two events A and B happen together can be found by

$$P(A \text{ and } B) = P(A)P(B \mid A)$$

Here $P(B \mid A)$ is the conditional probability that B occurs given the information that A occurs.

Conditional Probability

Conditional probability is used when information is available that might change the probability of an event.

Consider the following table and calculate the following probabilities:

- Person selected is a female
- Person selected is unemployed
- Person selected is unemployed if it is known that is female

The last question involves a conditional probability. Basically, the information given restricts the possible outcomes.

Example

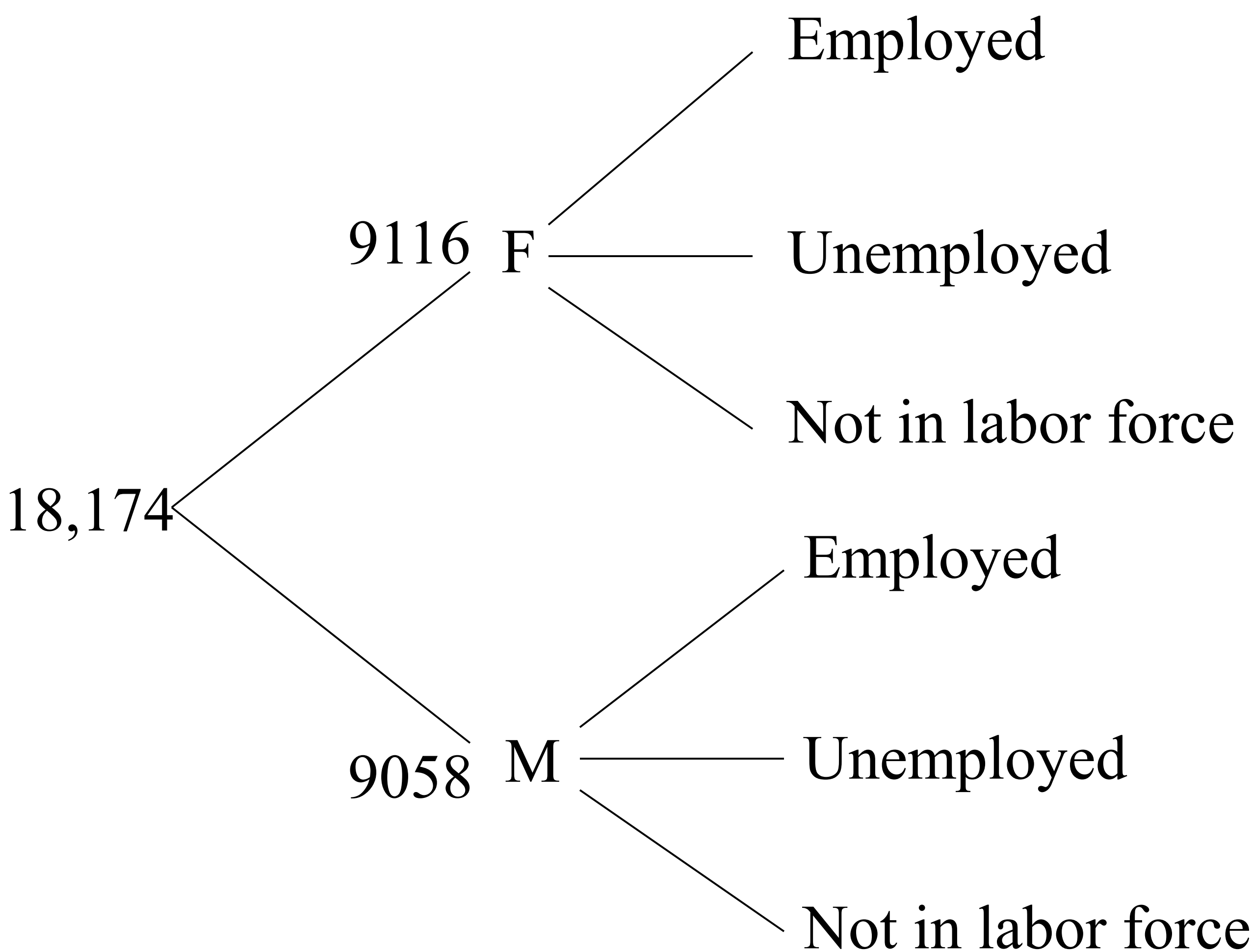
	Employed	Unemployed	Not in labor force	Total
Male	3927	520	4611	9058
Female	4313	446	4357	<u>9116</u>
Total	8240	<u>966</u>	8968	18174

P(Female) = P(Unemployed) =

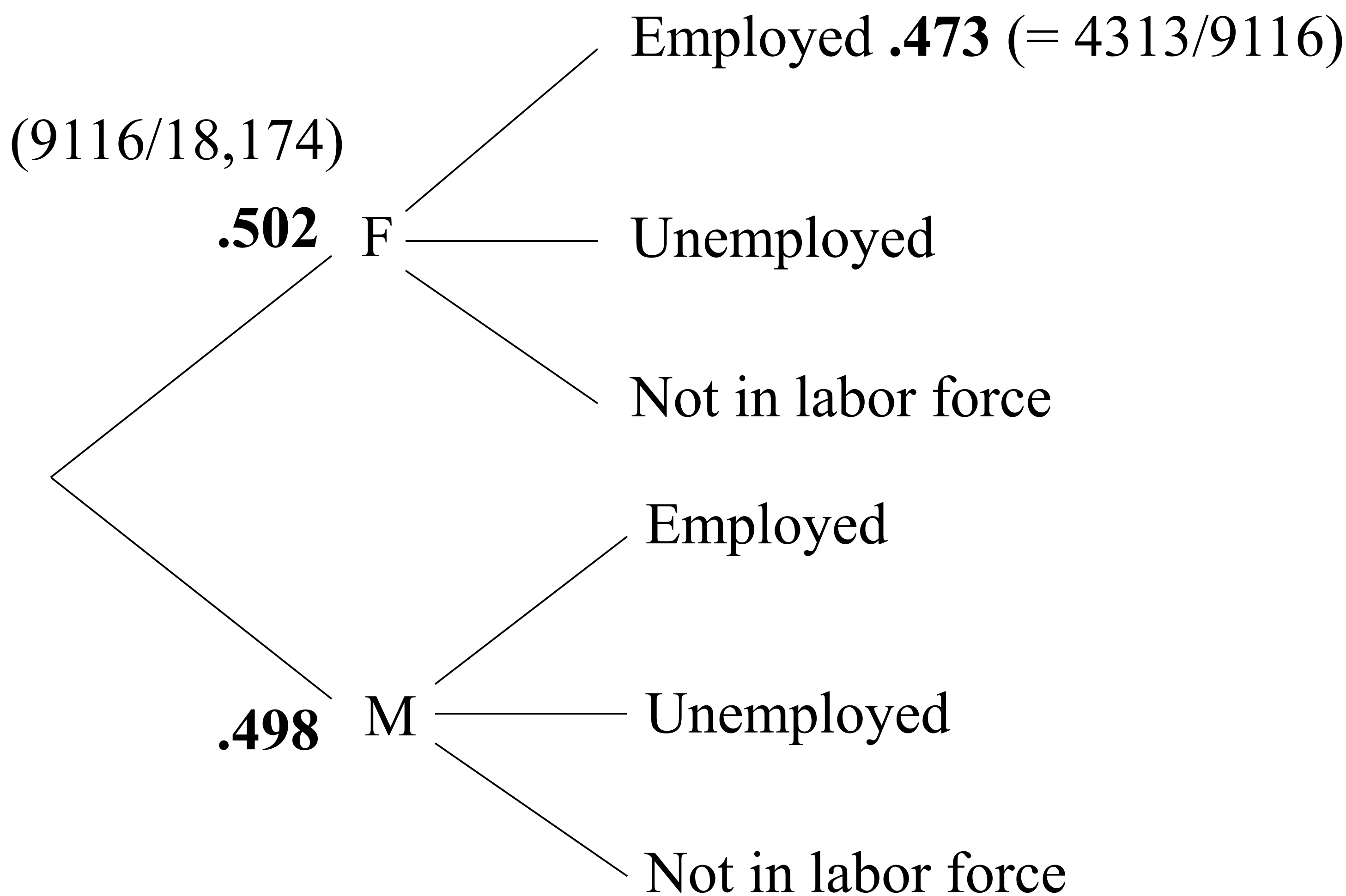
Person selected is unemployed and female: P(U and F) =

Suppose we want to calculate the frequency of an unemployed person such that the person is female : P(U|F) ?=

Tree Diagrams with Frequencies



Tree Diagrams with Proportions



(Proportions for the second branches are conditional on the first.)

Conditional Probability can be found directly from the data given rather than through the formula below.

DEFINITION OF CONDITIONAL PROBABILITY

When $P(A) > 0$, the **conditional probability** of B given A is

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

$$P(\text{Unemployed} \mid \text{Female}) = 446 / 9116 = \frac{(446 / 18{,}174)}{(9116 / 18{,}174)}$$

- Assume it is known that 2% of a population (that come for X-rays) have tuberculosis (T) $\underline{P(T) = .02}$

- If there is a 90% chance that the X-Ray of a tubercular person is positive… $\underline{P(Xray = +|T) = .90}$

- Calculate the probability that a person has a positive X-Ray and has tuberculosis $\underline{P(T \text{ and } (Xray = +)})$?

- P(A and B)= P(A) P(B|A)
- P(T and Xray+)=P(T) P(Xray+|T)=(0.02)(0.90)=0.018

MULTIPLICATION RULE FOR INDEPENDENT EVENTS

Two events A and B are **independent** if knowing that one occurs does not change the probability that the other occurs. If A and B are independent,

$$P(A \text{ and } B) = P(A)P(B)$$

which implies…

INDEPENDENT EVENTS

Two events A and B that both have positive probability are **independent** if

$$P(B \mid A) = P(B)$$

That is, knowing A happened does not affect the likelihood of B.