1 Random Variables

Q: What is Cauchy's least favorite expression?

A: "Got a moment?"

Q: $\times \times \quad \top\gamma\alpha\omega\exists \quad \underset{\smile}{\odot\odot}$

In this chapter we introduce random variables and their probability distributions. So far we have been concerned with random experiments, events, and their probabilities. However, no numerical values have been attached to the particular outcomes. Outcomes that can be associated with numbers lead us to the following definition of random variable.

A **random variable** is a variable whose numerical value is determined by the outcome of a random experiment.

We denote random variables by $X, Y, Z, \ldots$.

Example 1. Suppose a fair coin is tossed three times. We can define several random variables connected with this experiment. For example, X - number of heads, Y - difference between number of heads and number of tails, etc.

Each random variable can be fully described by its probability distribution.

Probability distribution of a random variable X is a table (assignment, rule, formula) which gives probabilities of particular realizations or sets of realizations.

Example 1 (contd.) For the random variable X possible realizations are 0 (no heads), 1 (exactly one head), 2 (exactly two heads), and 3 (all heads). Describing random variable X amounts to finding probabilities of all realizations. For instance, the realization $\{X = 2\}$ corresponds to the event $\{HHT, HTH, THH\}$. Thus, the probability of X taking value 2 is equal to the probability of the event $\{HHT, HTH, THH\}$ which is equal to 3/8. Finding probabilities for other outcomes we get the distribution of random variable[1] X which is given by the following table.

X	0	1	2	3
p	1/8	3/8	3/8	1/8

Most random variables of interest to us will be results of a random sampling. There is general classification of random variables which is based on the structure of realizations they can take. The random variables that take a finite or countable set possible discrete values are called **discrete random variables.** Random variable X from Example 1 is a discrete random variable. The second type of random variable takes values from (finite or infinite) interval. Results of measurements are usually modelled by continuous random variables.

[1] Term **random variable** is not the most fortunate. The better term would be **random function** or **random transformation** since X maps the sample space $\mathcal{S}$ to the set of real numbers.

2 Discrete Random Variables

Definition of Descrete Random Variable. Expectation + Variance.

Example: A jar contains two black and two white balls. Suppose that the balls have been thoroughly mixed and two are randomly selected from the jar.

(a) List all elementary outcomes for the experiment and assign appropriate probabilities to each. Make sure that the sum of the probabilities is 1.

(b) Let X be the number of white balls in the selection. Write the probability distribution for X.

Solution:

Apgar Score. At 1 min after birth and again at 5 min, each newborn child is given a numerical rating called an *Apgar score*. Possible values of this score are 0, 1, 2, $\cdots$, 9, and 10. A child's score is determined by five factors: muscle tone, skin color, respiratory effort, strength of heartbeat, and reflex, with a high score indicating a healthy infant. Let the random variable x denote the Apgar score of a randomly selected newborn infant at a particular hospital, and suppose that X has the given probability distribution.

X	0	1	2	3	4	5	6	7	8	9	10
Probability	.002	.001	.002	.005	.02	.04	.17	.38	.25	.12	.01

Definition of Expectation and Variance

Trick-or-treat When trick-or-treaters come to Peter Mueller's house, he blindfolds them and lets them draw a bank-note from a box. There are 10 bank-notes in the box, 5 one-dollar bills, 3 five-dollar bills ans 2 ten-dollar bills. Describe random variable X-the gain of trick-or-treater. Find EX and $VarX$.

Loaded Dice A loaded dice has the following probabilities of obtaining $X = k$ points.

$$P(X = k) = c \cdot k, \quad k = 1, 2, \ldots, 6,$$

where c is an unknown constant.

(a) Find the constant c, and write down the probability distribution for X.

(b) Find the probability of getting an odd number.

(c) If the die was rolled two times what is the probability of getting sum 12.

(d) Find EX and $VarX$.

(e) If the die was rolled 3 times, describe a random variable Y - the number of 6s.

Ice Cubes One of the statistics quoted by the *Wall Street Journal* is that the average number of ice-cubes placed in a cold drink is 3.2. Suppose that a study of a large number of drinks produced the following probability distribution for the number X of ice-cubes used per glass.

X	1	2	3	4	5	6
p	0.05	0.2	0.4	0.2	0.1	0.05

Find EX and $VarX$.

Result: $EX = 3.25, VarX = 1.3875$.

Expectation of Functions

Example: A sample of 10 temperature readings in Fahrenheit has a mean of 50 and a standard deviation of 9. What are the mean and the standard deviation in Celsius (centigrade)? (*Hint:* $\frac{F-32}{9} = \frac{C}{5}$)

Example: In a certain population the number of phones per household is modeled by a random variable X having the following distribution:

$X = \#$ phones:	1	2	3	4
$p(X)$:	.36	.48	.15	.01

Find the mean number of phones per household. [$EX = 1.81$].

Game. The gain in a game can be described by a random variable X given by the distribution:

X	-1	6
p	$9b$	b

(a) Find: b, $E(X)$, and $Var(X)$.
(b) Calculate the approximate value of

$$P(-35 \leq T \leq -25)$$

where T is the total gain of 100 people playing the game independently of each other.

Pizza Problem: When Giovanni of Cosa Nostra Pizzas makes pizza, the diameter is a random variable R with distribution

R	10	15	20
p	0.4	0.3	0.3

Find ER and EA, where A is the area of pizza. Show that $(ER)^2\pi \neq EA$. Expalin.

[ER=14.5, EA=714.7123, $(ER^2)\pi = 660.52$.]

Yet another game. A raffle ticket costs \$2.00. The firs prize is \$100, the second prize is \$50, the third is \$25. The probabilities of winning each prize are, for a ticket buyer, listed below.

Win	Prob.
\$100	.002
\$50	.015
\$25	.020

(a) What is the probability that a ticket buyer loses his \$2.00 ticket money?
(b) What is the expected gain for a ticket buyer? (Gain is win minus loss.)
(c) What is the conditional probability of a ticket purchaser winning first prize given that he wins anything?

2.1 Bernoulli and Binomial Random Variables

Bernoulli Def.

Binomial Definition

Fruits. A fruit grower claims that 2/3 of his peach crop has been contaminated by the medfly infestation. Find the probability that among 4 peaches inspected by this grower
(a) all 4 have been contaminated by the medfly;
(b) anywhere from 1 to 3 have been contaminated.

Physical aggression. Studies show that 40% of U.S. families use physical aggression to resolve the conflict. Suppose 20 families are selected at random. Find the probability that the number, X, that use physical aggression is

(a) Exactly 9;

(b) Between 7 and 10 inclusive;

Use Table I (Appendix A).

(c) Find the expectation (μ) and the variance (σ^2), of the random variable X.

VCRs. There are two different types of videocassette records (VCR's), VHS and Beta. Past records suggest that 80% of all those purchasing a VCR at a certain store choose a VHS type.

(a) What is the probability that at most 4 out of next 10 purchasers choose VHS.

(b) The store currently has 8 VCR's of each type in stock. What is the probability that each of next 10 customers can obtain his/her desired type from thr current stock?

Result: (a) $X \sim \mathcal{B}(10, 0.8)$; $P(X \leq 4) = ...$

(b) $P(2 \leq X \leq 8) = ...\square$

Quality Control. Large lots of incoming product aat a manufacturing plant are inspected for defectives by using a sampling scheme. Ten items are to be examined and the lot is to be rejected if two or more defectives are observed. If a lot contains exactly 5% defectived, what is the probability that the lot will be rejected? accepted?

Result: X - number of defectives found ammong the 10 examined. $X \sim \mathcal{B}(10, 0.05)$; $P(\text{accepted}) = P(X < 2) = P(X \leq 1) = 0.9139$. $\square$

Polygraph. A study by a federal agency concludes that polygraph (lie detector) tests given to people telling the truth have probability 0.2 of suggesting that the person is lying (Office of Technology Assessment, *Scientific Validity of Polygraph Testing*, Government Printing Office, Washington DC, 1983).

(a) A firm asks 12 job applicants about thefts from previous employers, using a polygraph to assess their truthfulness. Suppose that all 12 answer truthfully. What is the probability that the polygraph says at least three are deceptive?

(b) Among 12 truthful persons, what is the mean number who will be classified as deceptive?

(c) What is the standard deviation of this number?

Pot smoking. A nationwide survey of seniors by the University of Michigan reveals that almost 70% disapprove of daily pot smoking according to a report in *Parade,* September 14, 1980. If 12 seniors are selected at random and asked their opinion, find the probability that the number who disapprove of smoking pot daily is

(a) anywhere from 7 to 9;

(b) at most 5;

(c) not less than 8.

Tickets. From a box containing 50 tickets marked "yes" and 50 tickets marked "no," 10 tickets are selected at random with replacement. Find the following:

(a) P(all 10 tickets are marked "no")

(b) P(at least 1 ticket is marked "no")

(c) P(the first 5 tickets are marked "no" and the second 5 are marked "yes")

(d) P(exactly 5 tickets are marked "no") (*Hint:* the answer is *not* 1/2. This problem should be difficult for you.)

2.2 Poisson Distribution

Poisson

Purdue Exponent. The number of misspelled words on any page of the Purdue Exponent has a Poisson distribution with mean 2. The number of smudged words on any page has a Poisson distribution with mean 1. Evaluate the probability that there is at least one either misspelled or smudged word on the cover page of an Exponent.

Accidents. The number of accidents in an office during a six-week period is a **Poisson** random variable with $\lambda = 2$. Interest is in the probability that there are:

(a) exactly two accidents,

(b) at least one accident,

(c) 1 or fewer accidents,

in a six-week period.

Bottles. Two thousand bottles are transported from Durham to Greensboro. The probability that one bottle gets broken in the transport is 0.001. Assuming independence, find the probability that exactly two bottles get broken in the transport.

2.3 Hypergeometric Distribution

Hypergeometric

Committee. A committee of 3 has to be selected from the group of 10 students. Four students in the group are seniors. Describe random variable X - number of seniors in the committee.

Aces and Spades Describe r.v. X - number of (a) aces (b) ♠ in a five card poker-hand.

3 Continuous Random Variables

Cdf's, Densities, Quantiles, etc.

Simulations

4 Normal Random Variables

The references I have seen on this give the credit (or blame) to Quetelet. In the 19th century, it seems that everyone was using this distribution; the theoreticians because of the claim of the applied people that it generally held, and the applied people because of the belief that it had been proved theoretically. The Central Limit Theorem was in some sense proved early in that century, although filling in the gaps did not occur until the last decade. That it was easy to work with also helped, and there was the Gauss-Markov Theorem which showed that the procedures based on normality worked well even when it was not the case.

Anyhow, Quetelet claimed that this would be the distribution of the measurements of a "normal person". The French translates literally. There was no non-normal alternative considered, although some other distributions were known, and there was the Pearson and the Gram- Charlier families. (Herman Rubin)

Normal R.Vs

Example. A random variable X has normal $N(\mu,\ \sigma^2)$ distribution.
Find: (a) $p = P(1 \le X \le 3),\ if\ \mu = 1\ and\ \sigma^2 = 4.$
(b) $\sigma^2,\ if\ P(1 \le X \le 3) = 0.6828\ and\ \mu = 2.$
(c) $\mu,\ if\ P(X \ge 0) = 0.8\ and\ \sigma^2 = 4.$

Normal Practice. Let $X \sim \mathcal{N}(2, 3^2)$. Find:

(a) $P(0 < X < 5)$
(b) $P(X \ge 6)$
(c) $P(X < 0)$
(d) $P(X < x) = 0.9$, find x.

Solution: (a) $P(-0.67 < Z < 1) = \Phi(1) - \Phi(-0.67) = 0.8413 - 0.2514 = 0.5899.$
(b) $P(Z \ge 1.33) = 1 - \Phi(1.33) = 1 - 0.9082.$
(c) $P(Z < -0.67) = \Phi(-0.67) = 0.2514.$
(d) $P(Z > \frac{x-2}{3}) = 0.9.\quad \frac{x-2}{3} = 1.28,\ x = 3 \cdot 1.28 + 2 = 5.84.$

Grades. The results of a STA110 final may be modeled with the normal distribution with mean $\mu = 80$ and $\sigma = 8$. The policy of the exam is that only the upper 25 % of the scores are graded by an 'A'. Find the cut-point score.

Solution: $P(X < x) = 0.75; P(Z < \frac{x-80}{8}) < 0.75.$
$\frac{x-80}{8} = 0.675.$ $x = 80 + 8 \cdot 0.675 = 85.4$

Pulse Rate The pulse rate per minute of the adult male population between 18 and 25 years of age in the United States is known to have a normal distribution with a mean of 72 beats for minute and a standard deviation of 9.7.

(1) If the requirements for military service state that anyone with a pulse rate over 100 is medically unsuitable for service, what proportion of males between 18 and 25 years of age would be declared unfit because their pulse rates are too high.

(2) 80 % of the adult male population between 18 and 25 years of age in the United States has a pulse rate lower than Jim? What is Jim's pulse rate?

Cheese. The amount of moisture content (in pounds) in a 25-pound wheel of cheese is distributed as a normal $N(9, (0.9)^2)$ random variable. What is the probability that in a randomly chosen wheel the amount of moisture:

- Exceeds 10 pounds;
- Is between 8.1 and 9.9 pounds?
- Differs from the mean by no more than 1.8 pounds?
- Find the number x_0, so that the amount of moisture 99% of the wheels exceeds x_0.
- In what bounds (about the mean) the amount of moisture falls with the probability 0.95?
- If the amount of moisture is $N(9, \sigma^2)$, find σ so that

$$P(8 \leq X \leq 10) = 0.64.$$

Solution:

(1)

$$\begin{aligned} P(X > 10) &= 1 - P(X \leq 10) \\ &= 1 - P(Z \leq 1.11) = 1 - \Phi(1.11) \\ &= 1 - 0.8665 = 0.1335. \end{aligned}$$

(2)

$$\begin{aligned}
P(8.1 \leq X \leq 9.9) &= P(\frac{8.1-9}{0.9} \leq Z \leq \frac{9.9-9}{0.9}) \\
&= P(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) = 0.8413 - 0.1587 \\
&= 0.6826.
\end{aligned}$$

Note $1-\sigma$ rule.

(3)

$$\begin{aligned}
P(9-1.8 \leq X \leq 9+1.8) &= P(-2 \leq Z \leq 2) \\
&= \Phi(2) - \Phi(-2) = 0.9772 - 0.0228 = 0.9544.
\end{aligned}$$

Note $2-\sigma$ rule.

(4)

$$\begin{aligned}
0.99 = P(X > x_0) = 1 - P(X \geq x_0). \\
P(Z \leq \frac{x_0 - 9}{0.9}) = 0.01, \\
\Phi(\frac{x_0 - 9}{0.9}) = 0.01, \\
\frac{x_0 - 9}{0.9} = -2.33, \\
x_0 = 9 - 2.33\ 0.9 = 6.903.
\end{aligned}$$

(5)

$$\begin{aligned}
P(9-\delta \leq X \leq 9+\delta) = 0.95, \\
\Phi(\frac{\delta}{0.9}) = 0.975, \\
\frac{\delta}{0.9} = 1.96, \\
\delta = 1.96\ 0.9 = 1.764.
\end{aligned}$$

(6)

$$\begin{aligned}
P(\frac{-1}{\sigma} \leq Z \leq \frac{1}{\sigma}) = 0.64, \\
\Phi(\frac{1}{\sigma}) = 0.82, \\
\frac{1}{\sigma} = 0.915, \\
\sigma = 1.093.
\end{aligned}$$

4.1 Some inportant distributions that follow from the normal

t, F, chi

5 Simulations

Splus has capabilities to perform different probability calculations, random number generations and simulations.

A generic command is `sample.` For example

```
>sample(10)
 [1]  4  9 10  5  1  8  3  2  7  6
```

will give a random permutation of 10 numbers.

To get 14 numbers from 1 to 10 with replacement (without replacement it would be impossible) we say:

```
> sample(10,15,replace=T)
 [1]  2  9  9  4  7  7  6  3  5  4  1  4  7  5 10
```

Your favorite three states in US

```
> sample(state.name, 3)
[1] "Rhode Island" "Washington"   "Texas"
```

are in facet randomly chosen from the list of all US states. If we want to get a sample of size 30 from the discrete distribution

X	-1	0	2	5	10
p	0.2	0.1	0.3	0.3	0.1

the way to do it is:

```
 > sample( c(-1,0,2,5,10), 30, prob=c(0.2, 0.1, 0.3, 0.3, 0.1), replace=T)
 [1]   5   2   5   5   2   5   2   5   2   5   2   2  -1   5   5  -1   5   0   5
[20]  -1   5   5   5  10  -1   5  -1   5   5   5
```

We now explain how to use standard distributions in Splus. The basic command is the concatenation of a key letter and the name of distribution. Each distribution has arguments if its own.

The key letters are **p**, **d**, **q**, and **r**. They stand for *distribution* (area left of the point(s) in the argument), *probability* (ordinate of the density or probability mass function for the point(s) in argument, *quantiles* for the probability(ies) in the argument, and *random numbers.* The names of distributions are mnemonic and easy to memorize: **unif** for uniform, **norm** for normal, **binom** for binomial, **t** for Student's t, **f** for Fisher's F, **chisq** for χ^2, **pois** for Poisson, etc.

Examples:

Prefix p. Arguments for the prefix p are the number or vector `x`, and spacific distribution parameters. The result is (are) the area(s) under the density curve left of the point(s) `x`. If the distribution is discrete the result is total probability mass corresponding to realizations left of the point `x`. To simulate normal tables, say

```
> pnorm(1.25)
[1] 0.8943502
```

Even better, Splus will do standardization procedure for you. Let $X \sim \mathcal{N}(1, 4^2)$. Then $P(X \leq 6)$ is

```
>pnorm(6, mean=1, s=4)
[1] 0.8943502
```

For Student's t with 10 degrees of freedom and χ^2 with 15 degrees of freedom the area below the density curve above the interval [1,3], or simply $P(1 \leq X \leq 3)$ is

```
>  pt(3, df=10) - pt(1, df=10)
[1] 0.1637747
>  pchisq(3, df=15)-pchisq(1, df=15)
[1] 0.000401945
```

Prefix d. Arguments for the prefix d are the number or vector of numbers `x` and the specific density parameters. The result is the ordinate of the density (continuous case) or the probability mass at the point(s) (discrete case). To find all binomial $\mathcal{B}(10, 0.4)$ probabilities write

```
> dbinom(0:10, size=10, p=0.4)
 [1] 0.0060466176 0.0403107840 0.1209323520 0.2149908480 0.2508226560
 [6] 0.2006581248 0.1114767360 0.0424673280 0.0106168320 0.0015728640
[11] 0.0001048576
```

For example $P(X = 2) = 0.1209323520$. If we want to plt a density graph of χ^2 distribution with $n = 10$ degrees of freedom, the way to do it is:

```
>  x_seq(0,30, length=100)
>  y_dchisq(x, df=10)
>  plot(x,y, type="l")
```

Prefix q. Arguments for the prefix q are the probability or the vector of probabilities and the result(s) is (are) corresponding quantile(s). You may think that input is the area under the curve "left of a point" and the output is the point. Table 14.4 (Standard Normal Quantiles) may be reconstructed as

```
> qnorm(c(0.75,0.8,0.85,0.9,0.95,0.975, 0.98,0.99, 0.995,0.999))
 [1] 0.6744898 0.8416212 1.0364334 1.2815516 1.6448536 1.9599640 2.0537489
 [8] 2.3263479 2.5758293 3.0902323
```

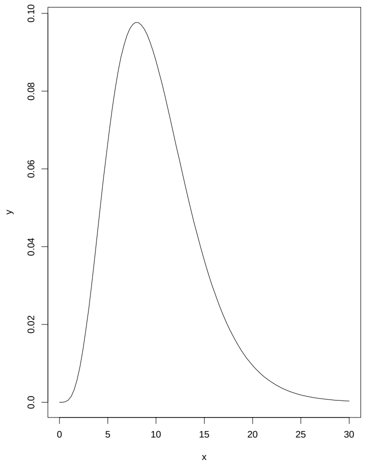

Figure 1: Density of χ^2 distribution with 10 degrees of freedom

Prefix r. Arguments for the prefix r are the number of random numbers wanted as well as the specific parameters of the distribution. For example, to generate 5 numbers from the Uniform (0,1) and 7 numbers from Uniform (1,6) distribution we write:

```
> runif(5)
[1] 0.9257951 0.4597539 0.4880530 0.5504434 0.0152659
> runif(7,min=1,max=6)
[1] 4.318238 5.609753 2.986946 1.734765 3.787045 1.507318 5.938067
```

For 18 random normal $\mathcal{N}(2, 4^2)$ random variables we write

```
> rnorm(18, mean=1, s=4)
 [1]  -7.6895603   3.9962352  -1.7901167  -1.9603230  11.8309521  -1.8583320
 [7]  -6.8676577   5.4059385  -2.9314005   3.8254438   7.4226048  -2.3012147
[13]   7.2775198   4.5356082  -4.4460672  -0.2172126  -0.3680125   2.8835324
```

Exercises:

1. Simulate and print 100 spins of a roulette (Numbers 1-40).

2. Simulate and print 100 rolls of a fair die.

3. Solve by using Splus:
$X \sim \mathcal{N}(-1, 3^2): \quad P(1 < X < 8)$,
$X \sim \chi^2_9: \quad P(6 < X < 10)$,
$X \sim t_{17}: \quad P(-2 < X < 1)$.
$X \sim t_3: \quad$ Find x_0 such that $\quad P(X > x_0) = 0.071$.
$X \sim F_{3,10}: \quad$ Find x_0 such that $\quad P(X < x_0) = 0.982$.

4. Plot the graph of $F_{4,12}$ distribution. (Fisher's F distribution with 4 numerator degrees of freedom and 12 denominator degrees of freedom.)

5. Length of life of many electric and electronic components is modelled by the exponential distribution. In Splus the exponential distribution is denoted by `exp`. The parameter of the distribution is λ.

(a) Plot the graph of the exponenyial distribution for the parameter $\lambda = 3$. (Make `x` grid to be nonnegative, say `seq(0,6,length=100)`.

(b) Generate 1000 random numbers and conclude that their average is close to 1/3. (Namely, for the exponential distribution expectation is $\frac{1}{\lambda}$.)

6 Exercises

Practice. Let Z have a standard normal distribution. Compute each of the following probabilities:

$$\begin{aligned} P(Z \geq -1), \\ P(-.1 \leq Z \leq .2), \\ P(-2.2 \leq Z \leq .8), \\ P(.3 \leq 1.5), \\ P(-5 \leq Z \leq 0), \text{ and} \\ P(Z \leq .6). \end{aligned}$$

More Practice. Let X have a normal distribution.

(a) If X has a mean of 10 and a standard deviation of 5, find each of the following robabilities:

$$\begin{aligned} P(X \geq 11), \\ P(X \leq 4), \\ P(X \geq 16), \\ P(7 \leq X \leq 17), \\ P(X \geq 8), \text{ and} \\ P(X \geq 0). \end{aligned}$$

(b) If X has a mean of 50 and a standard deviation of 100, find the following:

$$P(X \geq 150) \quad P(X \leq 0) \quad P(-1000 \leq X \leq 100)$$

(c) If X has a mean of -3 and a standard deviation of 8, find the following:

$$P(X \geq 5) \quad P(X \leq -7) \quad P(-15 \leq X \leq 1)$$

Tables. Let Z be standard normal. In each of the following cases find the number c which satisfies the equation:

$$\begin{aligned} P(Z \leq c) &= .95, \\ P(Z \geq c) &= .8, \\ P(-c \leq Z \leq c) &= .9, \text{ and} \\ P(-2c \leq Z \leq 2c) &= .9. \end{aligned}$$

Example. A stamping machine produces can tops whose diameters are normally distributed with a standard deviation of 0.01 inch. At what mean diameter should the machine be set so that no more than 4% of the can tops have diameters exceeding 3 inches?

Herrings. In a certain population of the herring Pomolobus aestivalis, the lengths of the individual fish follow a normal distribution. The mean length of the fish is 54 mm, and the standard deviation of the lengths is 4.5 mm. (Hildebrand, 1927).

(a) What percentage of the fish are between 54 and 63 mm long?

(b) What percentage of the fish are more than 58.5 mm long?

(c) Ten percent of the fish are longer than x. Find x.

Solvent. The weight of solvent in a steel drum is normally distributed, with mean $\mu = 352$ pounds and standard deviation $\sigma = 2.1$ pounds.

(a) For a random sample of six drums, what is the probability that the sample mean will not exceed 354 pounds?

(b) For a random sample of six drums, what is the range around μ within which $i\bar{X}$ will fall 90% of the time? (Use symmetrical limits around μ.)

Heights. If the heights of 300 students are normally distributed with mean 68.0 inches and standard deviation 3.0 inches, how many students have heights (a) greater than 72 inches, (b) less than or equal to 64 inches, (c) between 65 and 71 inches inclusive, (d) equal to 64 inches. Assume the measurements are recorded to the nearest inch.

Ans. (a) 20

Scores. The scores in a large class on a STAT 301T examination are approximately normally distributed with a mean of 75 and a standard deviation of 10. If 90% of the class is to pass, what should be the lowest passing score?

Pulse Rate. The pulse rate of one month old infants has a mean of 115 beats per minute and a standard deviation of 16 beats per minute.

(a) Explain shy the average pulse rate of a sample of 64 one month old infants is approximately normally distributed.

(b) Find the mean and the variance of the normal distribution in part (a).

(c) Find the probability that the average pulse rate of a sample of 64 will exceed 120.

7 Simulation

Simulation of different rvs in **S.**

8 Central Limit Theorem

Central Limit Theorem

Trial of the Pyx.[2] Since the early 13th century, coins struck by the Royal Mint in England have been evaluated for their metal content on a sample basis, in a ceremony called the Trial of the Pyx. This ceremony does not have much meaning anymore, but there was real money on the line back in the 1700s, because English coins were made of gold in those days. In 1799, for instance, the procedure went like this. One hundred gold coins called guineas were chosen at random from all of the coins made at the Mint that year, put in the Pyx (a ceremonial box), and weighed. The Master of the Mint, who was responsible for the quality of the coins, was allowed a margin of error, called the "remedy," which was set according to the manufacturing tolerances of the time.

In 1799 a guinea was supposed to weigh 128 grains (there are 360 grains in an ounce), so the 100 guineas in the Pyx should have weighed about 12800 grains. The remedy in those days was 1/400 of the expected amount, or 32 grains. If the actual weight of the coins in the Pyx differed from its expected value by more than the remedy on either the high or low side, the Master of the Mint was exposed to serious penalties. The British government had a vested interest in the coins not weighing too much, but the Master of the Mint had an incentive to make them weigh less than the standard, because he was allowed to keep the shortfall himself (as long as he was not caught by the Trial of the Pyx).

If the Master of the Mint is honest and manufactures guineas that weigh exactly 128 grains on average, with a SD of 1 grain, what is the chance that he will survive the Trial of the Pyx? To answer this question, first build a probability model, being explicit about the population and the sample.

If instead he sets things up so that the guineas weigh only 127.7 grains on average (with the same standard deviation of 1 grain), what is the chance now that he will survive the Trial? If he does survive, how much gold can he expect to pocket in an average year in which he produces 100,000 guineas? Give or take how much?

[2] Robert Gould `rgould@stat.ucla.edu`

Solution to Coin Questions 1

The population of interest is all coins minted in 1799, with mean 128 gr. and a standard deviation of 1 gr. From this population a simple random sample of 100 coins was selected for the Trial of the Pyx.

Let S be the sum of the weights of the 100 coins in the sample. The expected value of the sum, $E(S)$, is 100 times the population mean: 12800 gr. The standard error (SE) is given by the SD times the square-root of the sample size: 1*10 = 10 gr. Due to the **Central Limit Theorem**, the long run histogram of S is normal, centered at 12800 gr with variance equal to 100 (the square of the SE). The Master of the Mint survives the Trial of the Pyx only if the total weight S of the 100 coins in the sample weighs between (12800-32 and 12800+32) gr. We need to calculate $P(\text{survive}) = P(12768 < S < 12832)$. Normalize by subtracting the mean (12800) from all three quantities inside the probability and by dividing by the standard error. Using $Z = (S - E(S))/SE$ makes this probability equal to $P((12768 - 12800)/10 < Z < (12832 - 12800)/10) = P(-3.2 < Z < 3.2)$ which is approximately 99.9%.

Therefore, if the Master of the Mint is honest he is virtually certain to survive the Trial.

Solution to Coin Question 2

If the Master of the Mint sets things up so that the guineas weigh only 127.7 gr., then the population of interest remains the same as in the last question but with a mean of 127.7 gr. and SD still 1 gr.

The expected value of the sum S of the weights of the 100 coins in the sample is now $E(S) = n * E(S) = 12770gr$, with $SE(S) = 10$ gr. As before the long run histogram of S is normal, centered at 12770 gr and with variance equal to 100 (10 squared).

Then

$$\begin{aligned} P(\text{survive}) &= P(12768 < S < 12832) \\ &= P(\frac{12768 - 12770}{10} < Z < \frac{12832 - 12770}{10}) \\ &= P(-.2 < Z < 6.2) \end{aligned}$$

which is approximately 58 %. Therefore, his chances of surviving are better than 1 in 2.

However, if he manages to survive he will keep a lot of gold. It will take him $n * E(X) = 12,770,000$ gr to manufacture the 100,000 coins, give or take $\sqrt{n}SD = \sqrt{1000000} = 316$ gr. He will receive $100000 \cdot 128 = 12,800,000$ gr to manufacture them, so he will end up with 12,800,000 - 12,770,000 = 30,000 gr, give or take about 316 gr.

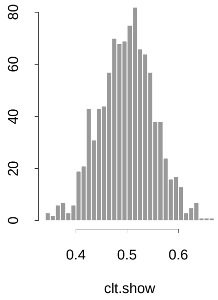

Figure 2: $N = 1000$ and $k = 1, 2, 3$, and 6.

Computer Example: In the following example we generate $N * k$ random numbers. Our choice is non-normal uniform distribution: `runif(N*k)`. These random numbers we format as a matrix (table) of dimension $N \times k$ (N rows, and k columns). Next we apply averaging operation on rows of a matrix. The result is a vector of lenght N. Each coordinate is a mean of the corresponding row. With $N = 1000$ and $k = 6$, the **S** statements are:

```
> clt.show.6_apply(matrix(runif(1000*6),ncol=6),1,mean)
> hist(clt.show.6, nclass=25)
```

Figure 2 shows histograms of averages of 1 , 2, 3, and 6 numbers.

Example: When you roll a fair die once you expect 3.5 points with a standard deviation of 1.708. (Variance 2.918).

Estimate the probability that in 1000 rollings the total sum will be between 3500 and 3600 points.

Gain in a long run. The gain in a game is described by

X	-1	0	5
p	0.2	0.2	0.6

(a) Would you play the game? Explain.

(b) If you play the game 100 times what are the chances of ending with a negative balance?

[Sol. $EX = 2.8, VarX = 7.36$, $\sigma X = 2.71, S = X_1 + \ldots + X_{100}$, *$S$ is approx. normal (CLT),* $ES = 280, VarS = 736, \sigma S = 27.1$, $P(S < 0) = P(Z < \frac{0-280}{27.1}) = P(Z < -10.33) = 0.$*]*

Sampling Distributions

Teaching Methods. A recent study in the *College Student Journal* (Sheehan et al. 1992) investigated differences in traditional and non-traditional students, where non-traditional students are generally defined as those 25 years or older. Suppose that a random sample of $n = 145$ non-traditional students is selected from 1993 population of non-traditional students and GPA of each student is determined. Assume that the population mean and the standard deviation for GPA of all non-traditional students is $\mu = 3.5$ and $\sigma = 0.5$. Then $\bar{X}$, the sample mean, will be approximately normally distributed because of ______ .

(a) Calculate the mean and variance of $\bar{X}$.

(b) What is the approximate probability that the non-traditional student sample has a mean GPA between 3.45 and 3.55?

(c) What would the answer for (b) be if n was 290.

Environment. Electric power plants that use water for cooling their condensers sometime discharge heated water into rivers, lakes, or oceans. It is known that water heated above certain temperatures has a detrimental effect on the plant and animal life in the water. Suppose it is known that the increased temperature of the heated water discharged by a certain power plant on any given day has a distribution with a mean of $5°C$ and a standard deviation of $0.5°C$.

(a) For 100 randomly selected days, what s the approximate probability that the **average** increase if temperature of the discharged water is less than $4.95°C$?

(b) Why can you use the normal distribution?

Sheriff. The Sherrif of Nottingham is taking some badly needed archery practice. He shoots at the target 100 times, and has a probability of hitting the target of 0.6 with each shot. The shots are independent. What is the chance that he will hit the target at least 70 times?

SAT. Scholastic aptitude test scores of college-bound high school seniors have a mean of $\mu = 475$ and a standard deviation of $\sigma = 90$.

a) What are the mean and standard deviation of the sampling distribution of $\bar{X}$ for a random sample of size of $n = 225$?
b) Find an approximation of the the probability that $\bar{X}$ exceeds 487.
c) What is the probability that an individual score will exceed 487, if we assume that the individual scores are normally distributed?

200 tosses. Find the probability that 200 tosses of a fair coin will result in (a) between 80 and 120 heads inclusive, (b) less than 90 heads, (c) less than 85 heads or more than 115 heads.

Ans: (Without adjustments) $np = 200 \cdot 0.5 = 100, \sigma^2 = npq = 50.$
(a) $P(\frac{80-100}{7.071} \leq Z \leq \frac{120-100}{7.071}\ \Phi(2.83) - \Phi(-2.83) = 0.9977 - 0023 = 0.9954$
With adjustments: 0.9962.

Bolts. A machine produces bolts which are 10% defective. Find the probability that in a random sample of 400 bolts produced by this machine (a) at most 30, (b) between 30 and 50, (c) 55 or more, of the bolts will be defective.

(dist.tables) Find 0.95-quantile of
(a) $N(2, 4^2)$,
(b) t_5, (Notation $t_{0.05,5}$)
(c) χ^2_{10}, (Notation $\chi^2_{0.95,10}$)
(d) $F_{3,5}$,
distribution.

To be moved to probability chapter. An electronic component consists of three parts. Each part has probability 0.9 of performing satisfactorily. The component fails if 2 or more of the parts do not perform satisfactorily. Assuming that the satisfactory performance of one part in no way depends on the performance of the other parts, determine the probability that the component does not perform satisfactorily.

Result: $(0.1)^3 + 3 \cdot (0.1)^2 \cdot 0.9 = 0.028 \quad \square$.

9 Exercises

Sales by Phone. On the average, 2 out of 10 people will respond favorably to a certain telephone sales pitch. If 20 people are called, letting S = the number who respond favorably, find:
(a) $P(S \geq 3)$
(b) $P(S < 6)$

Multiple Choice and Lazy Student. A professor gives a 100-question multiple-choice final exam. Each question has 4 choices. In order to pass, a student has to obtain at least 30 correct answers. A lazy student decides to guess at random on each question. What is the probability that the student passes the exam?

Some other Stat course. A statistics professor believes that the students in his recent introductory course were were exceptional. In the past about half the students scored higher than 70 on his final exam, but in his most recent class of 50 students, 35 students achieved scores above 70. If the students were actually similar to students in previouus classes, what would be the probability that at least 35 students would achieve scores above 70?

Morse code. Transmitting by Morse code cosists of sending a sequence of electronic signals of two types: short (represented by $\cdot$) and long (-). These signals are received by another station, pieced together, and translated into a message. Suppose the receiver correctly interprets a signal with probability .95.

(a) What is the probability that the transmission of 900 signals there are fewer than 100 errors?

(b) Suppose the system is modified so that each signal is transmitted three times. Then the receiver interprets any of the sequences $---$ or $--\cdot$ or $-\cdot-$ or $\cdot--$ as $-$. It interprets any of the signals $\cdot\cdot\cdot$ or $\cdot\cdot-$ or $\cdot-\cdot$ or $-\cdot\cdot$ as $\cdot$. Under the modified system what is the probability of incorrectly interpreting the signal sent? What is the probability that the number of errors in 900 signals transmitted is fewer than 100?

Sweepstakes. Jason went to his mailbox one afternoon and discovered that he had been chosen as semifinalist in the Publishing Clearing House Sweepstakes. This entitled him to enter the grand prize drawing for \$ 1 million dollars!!! Being a bit sceptical, Jason sat down with his calculator and statistics knowledge to figure out how much money he could expect to win if he played. The letter provided the following prize breakdown:

1	Supremo Grand Prize	\$ 1 million
5	Spectacular First Prizes	\$ 1,000
25	Swell Second Prizes	\$ 100
50	Thrilling Third Prizes	\$ 10

If 250,000 other people have also been selected as semifinalists, how much money could Jason approximately expect to win if he entered. What is the variance of his possible gain. Ignore the cost of a stamp.

Budbowl. Immediately after watching the Superbowl a viewer poll was taken. It was found that 40 % of the viewers could not recall the winner of the Budbowl and another half of those could not remember the final score of the game. 60% of the viewers had total recall.

If 10 people are randomly selected from a large number viewers of the game, what is the probability that at least 8 of them cannot recall who won the Budbowl? (by the way the answers are Bud and 55-10).

```
> 1-pbinom(7, size=10, prob=0.4)
[1] 0.01229455
or
> dbinom(8,10,0.4)+dbinom(9,10,0.4)+dbinom(10,10,0.4)
[1] 0.01229455
```

Clearance. The height if trucks on the interstate I-40 is approximately normally distributed with mean 10 ft and standard deviation 1.5 ft. Design the minimal clearance D, at an overpass under construction so that the probability that the truck will clear is 0.999.

```
> qnorm(0.999, mean=10, s=1.5)
[1] 14.63535
```

Bioassay. In a bioassay experiment, mice are exposed for two hours per day to 2.0 ppm NO_2, to simulate exposure to fossil-fuel combustion products. The number X of lesions (sore spots) is counted on each exposed mouse. We will regard X as a random variable; a magic birdie has informed us that the probability distribution of X is:

$$P(X = 0) = 0.5 \quad P(X = 1) = 0.2 \quad P(X = 2) = 0.1 \quad P(X = 3) = 0.2.$$

In a sample of 100 mice, about how many total lesions would you expect to see?

Stat Cola. The mean content of a new canned Stat Cola is $\mu = 12$ oz. Assume the contents are normally distributed with known $\sigma = 0.6$ oz. Suppose a random sample of $n = 8$ of these canned sodas is to be selected. What is the probability that the mean content, $\bar{X}$, of the sodas obtained will be within 0.5 ounces of the true mean content of $\mu = 12$ oz.

```
> pnorm(12.5,mean=12,s=0.6/sqrt(8))-pnorm(11.5,mean=12,s=0.6/sqrt(8))
[1] 0.9815779
```

Random Choice. Use the random digits below to choose a simple random sample of 4 from the following list of employees of a small company.

Agarwal
Berger
Davis
Fuest
Garcia
Harrison
Hendrix
Hornblower
Jones
Lynch
Mercury
Munster
Wilson
Zappa

Random Digits: 12163 51409 11869 22903 06288 90309 96170

Loaded Die. Let X be the result of rolling a loaded die with probability distribution given below.

x_i	1	2	3	4	5	6
p_i	$\frac{3}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$

a. What is μ_X.

b. The law of large numbers says that after 1,000 rolls, the proportion of 6's should be very close to $\frac{3}{12} = 0.25$. True or False.

c. Let X_1 be the first roll of the die, let X_2 be the second, and let Y be the sum of two rolls $Y = X_1 + X_2$. If the variance of X_1 is 3.9, what is the mean and variance of Y?

Construct. Construct (give the probability distribution) an example of a discrete random variable that satisfies the following requirements:
- it takes only three different values with positive probabilities;
- it has a mean equal to 0;
- it has a variance equal to 2.

Archery Practice. The Sheriff of Nottingham is taking some badly needed archery practice. He shoots at the target 3 times and has a probability of hitting the target of 0.6 with each shot. The shots are independent. Describe the random variable: X=number of hits in 3 trials.

Macrolepiota Procera. The size of mushroom caps varies. While many species of *Marasmius* and *Collybia* are only 12-20 mm (1/2-3/4 in) in diameter, some fungi are nearly 200mm (8 in) across. The cap diameter of parasol mushroom (*Macrolepiota Procera*) is a Normal random variable with parameters $\mu = 230mm$ and $\sigma = 25mm$.
(a) What proportion of parasol caps has a diameter between 200 and 250 mm.
(b) 5% of parasol caps are larger than x_0 in diameter. Find x_0.

Stat Cola versus Prob Cola. In the Statland there are only two types of Cola: Prob Cola and Stat Cola. Only 20 % of cola drinkers prefer Prob Cola. Let X be the number of Stat Cola drinkers out of 4 randomly selected cola drinkers. Construct the probability distribution for the random variable X

```
> dbinom(c(0,1,2,3,4), 4, 0.8)
[1] 0.0016 0.0256 0.1536 0.4096 0.4096
```

Golfer. If it is assumed that a golfer will hit a drive into a sand trap 18% of the time, what is the probability that the player will hit the ball into a sand trap
(a) exactly three times out of first six holes;
(b) between 180 and 185 times (inclusive) out of first 1000 holes.
[Sol. (a) 0.0643, (b) With corrections: $\Phi(0.44) - \Phi(-0.041)$.]

Wild Thing. Suppose the impossible happens and the next World Series is matchup between the Cleveland Indians of the American League East and the Chicago Cubs of the National League West. If the Indians are slightly better team and have a 55% of prevailing on any given day, what is the probability that the Tribe wins the series in six games. (Assume the games are independent events)

[Sol. 0.18]

Aggression. In the five fourth-grade classes at a particular elementary school with large number of students, 10% of the students are considered aggressive based on peer and teacher reports of their behavior. Suppose 3 fourth-graders are selected at random.

(i) Describe (write down the probability distribution) the random variable X-number of aggressive children in the sample.

(ii) What is $P(X \geq 2)$.

(iii) What is EX and $Var(X)$.

Spontaneous Recovery. Twenty percent of individuals who seek psychotherapy will exhibit a return to normal personality irrespective of whether or not they receive psychotherapy (a phenomenon called *spontaneous recovery*).

(a) If a researcher studies $n = 7$ people who seek psychotherapy, what is the probability that at most 2 of them will exhibit spontaneous recovery.

(b) If the researcher studies $n = 200$ people who seek psychotherapy, what is an approximation to the probability that at most 49 of them will exhibit spontaneous recovery.

Sea Urchins. In a laboratory experiment, researchers at Barry University, (Miami Shores, Florida) studied the rate at which sea urchins ingested turtle grass (*Florida Scientist*, Summer/Autumn 1991). The urchins starved for 48 hours, were fed $5cm$ blades of green turtle grass. The mean ingestion time was found to be 2.83 hours and the standard deviation .79 hour. Assume that green turtle grass ingestion time for the sea urchins has an approximately normal distribution.

Find the probability that a sea urchin will require between 2.3 and 4 hours to ingest a 5 cm blade of green turtle grass.

Figure 3: Hi! My name is Sally Sea Urchin.

Mad Cow Desease. The average number of dairy cows suffering from BSE *Bovine spongiform encephalopathy* or " Mad Cow Desease" in a herd of 1000 cows in Cambridgeshire, England is estimated to be 4.5. Using the Poisson approximation, evaluate the probability that exactly two cows suffer from BSE in a herd of dairy 200 cows from that region.

Light Bulbs. Electric light bulbs bought for lighting an outdoor sports ground have a mean life of 3000 hours and a standard deviation of 340 hours. Assuming a normal distribution for the lifetime of bulbs, what proportion of light bulbs last longer than 2800 hours? If it is more economical to replace all of the bulbs when 20% of them have burned out than to change them as needed, after how many hours should the bulbs have to be changed?

BSE. Proportion of cattle developing BSE *Bovine spongiform encephalopathy* (Mad Cow Desease) at specific ages in the UK is given in the following table.

Three independent cases are selected. What is the probability that in at list one case the infested cow survived at most 6 years.

(a) Find the mean and variance of the random variable X- Age of death.

(b) Find the skewness and kurtosis of X.

10 Bayes Theorem and Bayesian Inference

Recall that multiplication rule claims:

$$P(AH) = P(A)P(H|A) = P(H)P(A|H).$$

This simple identity is the essence of **Bayes Theorem:**

$$P(H|A) = \frac{P(A|H)P(H)}{P(A)}.$$

Assume that the whole sample space S is partitioned by events $H_1, H_2, \ldots, H_n$. Events H_i we will cal **hypotheses** and they satisfy:

$$\begin{aligned} H_1 \cup H_2 \cup \ldots \cup H_n &= S \\ H_i \cap H_j &= \emptyset, i \neq j. \end{aligned}$$

If we are interested in probability of event A, then **formula of total probability** gives

$$P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + \ldots + P(A|H_n)P(H_n).$$

1. A new test has been devised for detecting a particular type of cancer. If the test is applied to the person who actually has that type of cancer, the probability that person will have a positive reaction is 0.995. If applied to a person who does not have this type of cancer, the probability that the person will have a (false) positive reaction is 0.012.

Suppose that test is given to patients at high risk of having cancer, and that one person, out of 50 in this group actually has this type of cancer.

(i) What is the probability that the randomly selected person from the group tests *positive?*

(ii) If randomly selected person tests *positive*, what is the probability that he/she **does not have** this type of cancer.

2. Jeremy, an enthusiastic Duke student, poses a statistical model for his scores on standard IQ tests. He thinks that, in general, his scores are normally distributed with unknown mean θ and the variance 80. Prior (and expert) opinion is that the true IQ of Duke students, θ, is a normal random variable, with mean 110 and the variance 120. Jeremy took the test and scored 98.

Estimate his true IQ θ in a Bayesian manner.

9. A student answers a multiple choice examination question that has 4 possible answers. Suppose that the probability that the student knows the answer to the question is 0.80 and the probability that the student guesses is 0.20. If student guesses, probability of correct answer is 0.25.

(i) What is the probability that the fixed question is answered correctly?

(ii) If it is answered correctly what is the probability that the student really knew the correct answer.

10. Factory has three types of machines producing an item. Probabilities that the item is I quality f it is produced on i-th machine are given in the following table:

machine	probability of I quality
1	0.8
2	0.7
3	0.9

The total production is done 30% on type I machine, 50% on type II, and 20% on type III.

One item is selected at random from the production.

(i) What is the probability that it is of I quality?

(ii) If it is of first quality, what is the probability that it was produced on the machine I?

11. One out of 1000 coins has two tails. The coin is selected at random out of these 1000 coins and flipped 5 times. If tails appeared all 5 times, what is the probability that the selected coin was 'two-tailed'?

19. In the city Kokomo, IN, 50% are conservatives, 30% are liberals and 20% are independents.

Records show that in a particular election 82% of conservatives voted, 65% of liberals voted and 50% of independents voted.

If the person from the city is selected at random and it is learned that he/she did not vote, what is the probability that the person is liberal?

1.[10 pts] A new test has been devised for detecting a particular type of cancer. If the test is applied to the person who actually has that type of cancer, the probability that person will have a positive reaction is 0.995. If applied to a person who does not have this type of cancer, the probability that the person will have a (false) positive reaction is 0.012.

Suppose that test is given to patients at high risk of having cancer, and that one person, out of 50 in this group actually has this type of cancer.

(i) What is the probability that the randomly selected person from the group tests *positive?*

(ii) If randomly selected person tests *positive*, what is the probability that he/she **does not have** this type of cancer.

A federalist paper resolved! Suppose that a work, the author of which is known to be either Madson or Hamilton, contains a certain key phrase. Suppose further that this phrase occurs in 60% of the papers known to have been written by Madison, but in only 20% of those by Hamilton. Finally, suppose a historian gives subjective probability .3 to the event that the author is Madison (and consequently .7 to the complementary event that the author is Hamilton). Compute the historian's posterior probability for the event that the author is Madison.

Transylvania. In a small village in Transylvania 15% of the population are vampires, 20% are ghosts, and 65% are ordinary people. Vampires never tell the truth, ghosts tell the truth 37% of the time, and ordinary people tell the truth 95% of the time. It is impossible to tell apart vampires, ghosts and ordinary people by the way they look (between 6:00am and 11:59:59pm). You are introduced to a gentleman from the village.

(a) What is the probability that he did not give you his real name?

(b) Just at 11:59:59pm you realize that the gentleman lied about his name. What is the probability that your companion is a vampire or ghost?

1. Robin Hood had finally been caught by the sheriff of Nottingham and was scheduled for execution. Because Robin was a popular her and the sheriff wanted to seem generous, he offered Robin a chance to go free: "Here is box with 8 black balls and 2 white balls. You will pick one ball from the box while blindfolded, and if you pick a white ball you go free. If you pick a black ball, you die." Robin, who had taken a statistics course, proposed instead that he (Robin) be allowed to sort the balls into two boxes. The sheriff would then choose one of the two boxes at random and Robin would select from that box while blindfolded. The sheriff thought that Robin's suggestion would make no difference (the odds that Robin would go free would still be only $\frac{2}{10} = 0.20$). Hence he agreed to Robin's proposal. The sheriff should have taken a statistics course. Below are three ways in which Robin could arrange the 8 black balls and 2 white balls in the two boxes. In each case, find the probability that Robin goes free (picks a white ball).

(a) | ○ ○|| ● ● ● ● ● ● ● ●|.
(b) | ○ ●|| ○ ● ● ● ● ● ● ●|.
(c) | ○ || ○ ● ● ● ● ● ● ● ● |.

College Entrance Test. Because of the role of college aptitude test scores in college entrance decision, there are minicourses that purport to teach students how to take these tests. A particular aptitude test has been found to produce scores that are normally distributed, with mean θ and standard deviation 80. If the minicourse directed at this test is effective (on average), the mean score θ of students who take the course is larger than 500; otherwise it is not. We want to test

$$H_0 : \theta \leq 500 \quad \text{versus} \quad H_1 : \theta > 500,$$

and our prior for θ is $N(520, 20^2)$.

1. Find the prior probabilities of hypotheses H_0 and H_1, π_0 and π_1.

2. If 50 observations gives the mean 513, find the posterior probabilities of hypotheses H_0 and H_1, p_0 and p_1, and make the decision.

3. What is 90% credible set for θ.

Solution: XXXXXXXXXXX)

Two Masked Robbers. Two masked robbers try to rob a crowded bank during the lunch hour but the teller presses a button that sets off an alarm and locks the front door. The robbers realizing they are trapped, throw away their masks and disappear into the chaotic crowd. Confronted with 40 people claiming they are innocent, the police gives everyone a lie detector test. Suppose that guilty people are detected with probability 0.85 and innocent people appear to be guilty with probability 0.08. What is the probability that Mr. Smith was one of the robbers given that the lie detector says he is?

Guessing Subject in an experiment are told that either a red or a green light will flash. Each subject is to guess which light will flash. The subject is told that the probability of a red light is 0.7, independent of guesses. Assume that the subject is a probability matcher- that is , guesses red with probability .70 and green with probability .30.

(a) What is the probability that the subject guesses correctly?

(b) Given that a subject guesses correctly, what is the probability that the light flashed red?

Toothpaste. A research study on fluoride in toothpaste was conducted using Colgate's *MFP* formula, and a leading stannous fluoride (*SF*) toothpaste. Data[3] on the number of new cavities over three year period are summarized as follows:

	n	Mean	s.d.
MFP	208	19.98	10.60
SF	201	22.39	11.96

The mean difference, X, in number of new cavities between the *MFP* and *SF* samples is modeled by $N(\theta, 1.12^2)$ The null hypothesis is $\theta \leq 0$. Assume that the prior for θ is $N(4, 2^2)$.

(1) Find the posterior distribution for θ given the observed mean difference $X = 2.41$.

(2) Test the hypothesis H_0.

(3) Find 95% credible set for θ. Compare the credible set with the 95% confidence interval.

[3] Frankl, S. and Alman, J. *J. Oral. Therapeutics Pharmacol. 4 (1968), 443-449.*

Bayes and bats. By careful examination of sound and film records it is possible to measure the distance at which a bat first detects an insect. The measurements are modeled by normal distribution $N(\theta, 10^2)$, where θ is the unknown mean distance (in cm).

Experts believe that the prior suitably expressing uncertainty about θ is $\theta \sim N(50, 10^2)$.

Three measurements are obtained: 62, 52, and 68.

(a) Find the posterior distribution of θ given the observations.

(b) Test the hypothesis H_0 that $\theta \geq 50$ in Bayesian fashion.

(c) What is the 95% credible set for θ.

Crystal. Chrystal (1891) wrote: *No one would say that, if you simply put two white balls in a bag containing one of unknown color, equally likely to be white or black, that this action raised the odds that the unknown ball is white ftro even to 3:1.* If we draw the 3 balls and the two first are white, is Chrystal's argument still valid? (Note: He used this argument to reject the Bayes rule.)

College Entrance Test Again. Because of the role of college aptitude test scores in college entrance decision, there are minicourses that purport to teach students how to take these tests. A particular aptitude test has been found to produce scores that are normally distributed, with mean θ and standard deviation 60. If the minicourse directed at this test is effective (on average), the mean score θ of students who take the course is larger than 500; otherwise it is not. We want to test

$$H_0 : 480 \leq \theta \leq 520 \quad \text{versus} \quad H_1 : \theta \text{ not in } [480, 520],$$

and our prior for θ is $N(520, 30^2)$.

- (i) If 25 observations gives the mean 510, perform the test and make the decision.
- (ii) Find 96% credible set for θ. Compare the obtained credible set with the frequentist 96% confidence interval for the unknown mean θ. Explain why credible sets tend to be shorter than the corresponding confidence intervals.

Weapons at Airports. A recent test conducted by the Federal Aviation Administration found that guards hired to screen passengers for weapons at the boarding gate had a very poor rate of detection. The detection rates for weapons carried by F.A.A. inspectors or placed in their carry-on-luggage averaged 80%, but their rates varied from 34% to 99% for the airports tested (*New York Times,* June 18, 1987). Suppose that in a particular city, airport **A** handles 50% of all airline traffic, while airports **B** and **C** handle 30% and 20%, respectively. The detection rates at the three airports are 0.9, 0.5, and 0.4, respectively. If the passenger at one of the airports is found to be carrying a weapon through the boarding gate, what is the probability that the passenger is not using the airport **C**?