Ruxin Liu

3.2 **11** A restaurant offers apple and blueberry pies and stocks an equal number of each kind of pie. Each day ten customers request pie. They choose, with equal probabilities, one of the two kinds of pie. How many pieces of each kind of pie should the owner provide so that the probability is about .95 that each customer gets the pie of his or her own choice?

$n = 10$ $\quad P(\text{apple pie}) = \frac{1}{2}$

K: number of apple pie $\quad K \in [5, 10]$

Binomial distribution: $b(n, p, k) = b(10, 0.5, k)$

Calculate cumulative Probability for K

K	$P(X \leq x)$
5	0.623
6	0.828
7	0.945
8	0.989 ⟵ > 0.95

8 pieces of each kind is needed

3.2 **29** A drug is assumed to be effective with an unknown probability p. To estimate p the drug is given to n patients. It is found to be effective for m patients. The *method of maximum likelihood* for estimating p states that we should choose the value for p that gives the highest probability of getting what we got on the experiment. Assuming that the experiment can be considered as a Bernoulli trials process with probability p for success, show that the maximum likelihood estimate for p is the proportion m/n of successes.

Bernoulli trials $\longrightarrow$ Binomial distribution

$b(n, p, m) = \binom{n}{m} p^m (1-p)^{n-m}$

In order to find the maximum likelihood estimate,

$\frac{d\, b(n,p,m)}{dp} = 0 \qquad \frac{d \binom{n}{m} p^m (1-p)^{n-m}}{dp} = 0$, $\binom{n}{m}$ is constant

$p^m (n-m)(1-p)^{n-m-1} \cdot (-1) + m p^{m-1} (1-p)^{n-m} = 0$

$p^m (n-m)(1-p)^{n-m-1} = m p^{m-1} (1-p)^{n-m}$ $\qquad$ divide p^{m-1}

$p (n-m)(1-p)^{n-m-1} = m (1-p)^{n-m}$ $\qquad$ divide $(1-p)^{n-m-1}$

$p(n-m) = m(1-p)$

$pn - pm = m - mp \rightarrow pn = m \rightarrow \boxed{p = m/n}$

3.2

30 Recall that in the World Series the first team to win four games wins the series. The series can go at most seven games. Assume that the Red Sox and the Mets are playing the series. Assume that the Mets win each game with probability p. Fermat observed that even though the series might not go seven games, the probability that the Mets win the series is the same as the probability that they win four or more game in a series that was forced to go seven games no matter who wins the individual games.

(a) Using the program **PowerCurve** of Example 3.11 find the probability that the Mets win the series for the cases $p = .5$, $p = .6$, $p = .7$.

(b) Assume that the Mets have probability .6 of winning each game. Use the program **PowerCurve** to find a value of n so that, if the series goes to the first team to win more than half the games, the Mets will have a 95 percent chance of winning the series. Choose n as small as possible.

a)

```{r}
p <- 0.5
# Binomial distribution
dbinom(4, size = 7, prob = p) + dbinom(5, size = 7, prob = p) +
  dbinom(6, size = 7, prob = p) + dbinom(7, size = 7, prob = p)
```

win 4 or more games

```
[1] 0.5
```

```{r}
p <- 0.6
dbinom(4, size = 7, prob = p) + dbinom(5, size = 7, prob = p) +
  dbinom(6, size = 7, prob = p) + dbinom(7, size = 7, prob = p)
```

```
[1] 0.710208
```

```{r}
p <- 0.7
dbinom(4, size = 7, prob = p) + dbinom(5, size = 7, prob = p) +
  dbinom(6, size = 7, prob = p) + dbinom(7, size = 7, prob = p)
```

```
[1] 0.873964
```

b)

```
p <- 0.6
n = 10
dbinom(4, size = n, prob = p) + dbinom(5, size = n, prob = p) +
  dbinom(6, size = n, prob = p) + dbinom(7, size = n, prob = p) +
  dbinom(8, size = n, prob = p) + dbinom(9, size = n, prob = p) +
  dbinom(10, size = n, prob = p)
```

```
[1] 0.9452381
```

≈95% to win

n=10

5.2

1 Choose a number U from the unit interval $[0,1]$ with uniform distribution. Find the cumulative distribution and density for the random variables

(a) $Y = U + 2$.

(b) $Y = U^3$.

2 Choose a number U from the interval $[0,1]$ with uniform distribution. Find the cumulative distribution and density for the random variables

(a) $Y = 1/(U + 1)$.

(b) $Y = \log(U + 1)$.

Q1

a) $F(y) = P(Y \le y) = P(U+2 \le y) = P(U \le y-2) = \underline{y-2}$

$U \in [0,1]$ $\quad Y = U+2 \quad Y \in [0+2, 1+2]$ $\quad , \underline{Y \in [2,3]}$

$f(y) = \dfrac{d\,F(y)}{dy} = \underline{1} \quad , \underline{y \in [2,3]}$

b) $F(y) = P(Y \le y) = P(U^3 \le y) = P(U \le y^{\frac{1}{3}}) = \underline{y^{\frac{1}{3}}}$

$Y = U^3 \quad y \in [0^3, 1^3]$, $\underline{y \in [0,1]}$

$f(y) = \dfrac{d\,y^{\frac{1}{3}}}{dy} = \underline{\frac{1}{3} y^{-2/3} \quad , y \in [0,1]}$

$y \neq 0$

Q2

a) $F(y) = P(Y \le y) = (\frac{1}{U+1} \le y) = P(\frac{1}{y} \le U+1) = P(\frac{1}{y} - 1 \le U)$

$= 1 - P(U \le \frac{1}{y} - 1) = 1 - \frac{1}{y} + 1 = \underline{2 - \frac{1}{y}}$

$Y = \frac{1}{U+1} \quad Y \in [\frac{1}{0+1}, \frac{1}{1+1}]$, $\underline{y \in [\frac{1}{2}, 1]}$

$f(y) = \dfrac{d\,(2 - \frac{1}{y})}{dy} = \underline{\frac{1}{y^2} \quad , y \in [\frac{1}{2}, 1]}$

b) $F(Y) = P(Y \le y) = P(\log(U+1) \le y) = P(U+1 \le e^y)$

$= P(U \le e^y - 1) = \underline{e^y - 1}$

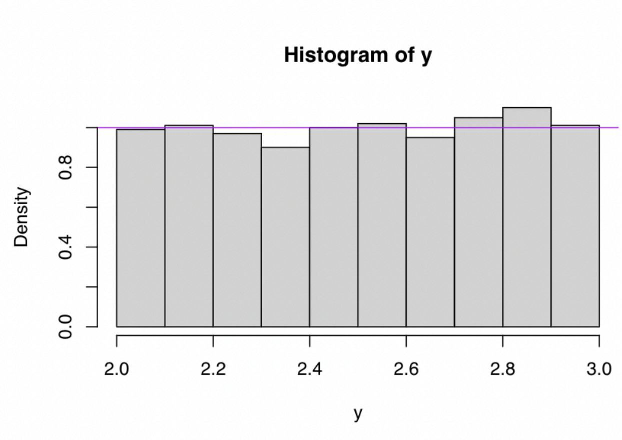
$Y = \log(U+1) \quad \underline{y \in [\log 1, \log 2]}$, $y \in [0, \log 2]$

$f(y) = \dfrac{d\,(e^y - 1)}{dy} = \underline{e^y \quad , y \in [0, \log 2]}$

5.2 **4** Suppose we know a random variable Y as a function of the uniform random variable U: $Y = \phi(U)$, and suppose we have calculated the cumulative distribution function $F_Y(y)$ and thence the density $f_Y(y)$. How can we check whether our answer is correct? An easy simulation provides the answer: Make a bar graph of $Y = \phi(rnd)$ and compare the result with the graph of $f_Y(y)$. These graphs should look similar. Check your answers to Exercises 1 and 2 by this method. Using R

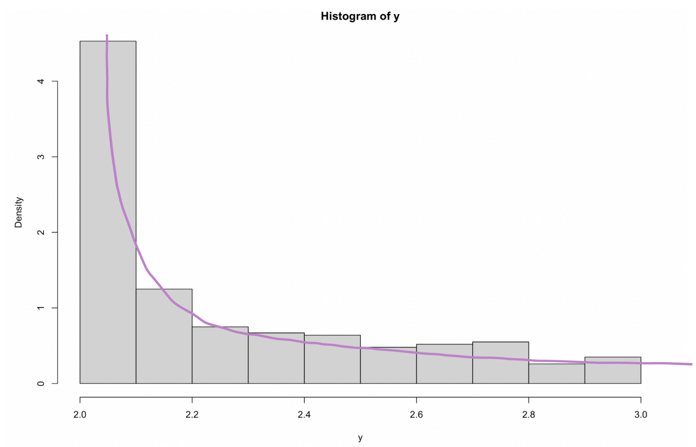
Q1 (a)

```
set.seed(19)
u <- runif(1000, 0, 1)
y <- u + 2
hist(y, freq = FALSE)
abline(1, 0, col = "purple")
```



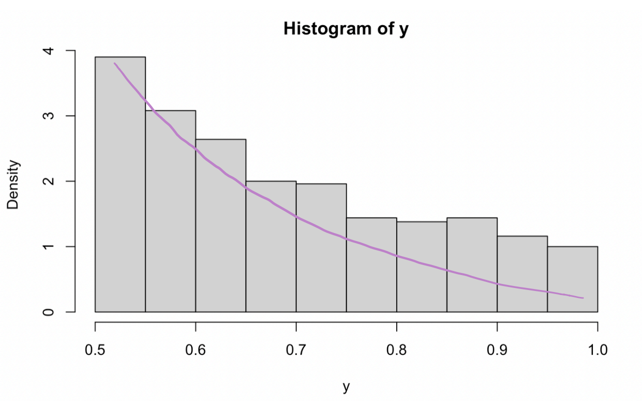
Q1 (b)

```
set.seed(19)
u <- runif(1000, 0, 1)
y <- u^3
curve(expr = 1 / 3 * x^(-2 / 3), from = 0, to = 1, col = "purple")
hist(y, freq = FALSE)
```



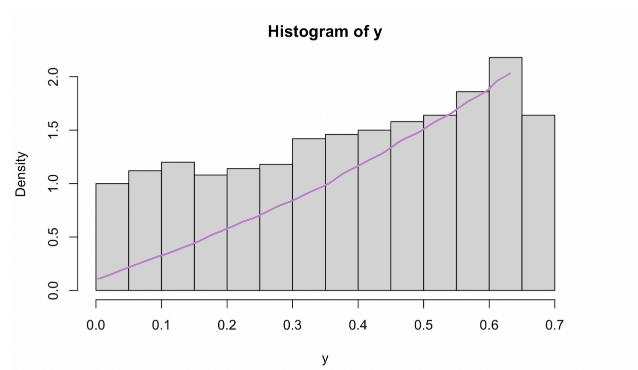
Q2 (a)

```
set.seed(19)
u <- runif(1000, 0, 1)
y <- 1 / (u + 1)
curve(expr = 1 / x^2, from = 0.5, to = 1, col = "purple")
hist(y, freq = FALSE)
```



Q2 (b)

```
set.seed(19)
u <- runif(1000, 0, 1)
y <- log(u + 1)
curve(expr = exp(x), from = 0, to = log(2), col = "purple")
hist(y, freq = FALSE)
```



5.2 **10** Let U, V be random numbers chosen independently from the interval $[0,1]$. Find the cumulative distribution and density for the random variables

(a) $Y = \max(U, V)$.

(b) $Y = \min(U, V)$.

Q10 a) $F(y) = P(Y<y) = P(\max(U,V) < y)$

$= P(U<y) \cdot P(V<y)$ ← independent

$= y \cdot y \quad = y^2$

$f(y) = \frac{dy^2}{dy} \quad = 2y \ , y \in [0,1]$

b) $F(y) = P(Y<y) = P(\min(U,V) < y)$

$= 1 - P(\min(U,V) > y)$

$= 1 - P(U>y) \cdot P(V>y)$

$= 1 - [P(U>y)]^2$

$= 1 - [1 - P(U<y)]^2$

$= 1 - (1-y)^2 \quad = 1 - 1 - y^2 \quad + 2y \quad = 2y - y^2$

$f(y) = \frac{d(2y-y^2)}{dy} \quad = 2 - 2y, \quad y \in [0,1]$

5.2 **26** Bridies' Bearing Works manufactures bearing shafts whose diameters are normally distributed with parameters $\mu = 1$, $\sigma = .002$. The buyer's specifications require these diameters to be $1.000 \pm .003$ cm. What fraction of the manufacturer's shafts are likely to be rejected? If the manufacturer improves her quality control, she can reduce the value of σ. What value of σ will ensure that no more than 1 percent of her shafts are likely to be rejected?

Q26 $P(1 - 0.003 < X < 1 + 0.003)$ ← P(not reject)

$= P\left(\frac{1 - 0.003 - 1}{0.002} < \frac{X - 1}{0.002} < \frac{1 + 0.003 - 1}{0.002}\right)$

$= P(-1.5 < Z < 1.5)$

$P(\text{reject}) = 1 - P(-1.5 < Z < 1.5)$

$= P(Z > 1.5) + P(Z < -1.5)$

$= 2\,\Phi(1.5) = 2 \times 0.0668 = 0.1336$

13.36% to be rejected

rejection < 1%

$2\Phi(Z) \leq 0.01$ $\Phi(Z) \leq 0.005$

In Z table, the 1st value ≤ 0.005 is 0.00494

$\Rightarrow Z = -2.58$

$\frac{-0.003}{\sigma} = -2.58$ $\Rightarrow \sigma = 0.0012$

For rejection rate no more than 1%, σ should be 0.0012

5.2 **28** (Ross[11]) An expert witness in a paternity suit testifies that the length (in days) of a pregnancy, from conception to delivery, is approximately normally distributed, with parameters $\mu = 270$, $\sigma = 10$. The defendant in the suit is able to prove that he was out of the country during the period from 290 to 240 days before the birth of the child. What is the probability that the defendant was in the country when the child was conceived?

Q28 $X \sim N(270, 10)$

P(defendant in the country)

$= 1 - P(240 \leq X \leq 290)$

$= 1 - P\left(\frac{240-270}{10} \leq Z \leq \frac{290-270}{10}\right)$

$= 1 - P(-3 \leq Z \leq 2)$

$= 1 - [\Phi(2) - \Phi(-3)]$

$= 1 - 0.97725 + 0.00135$

$= 0.024 = 2.4\%$

6.1

23 An insurance company has 1,000 policies on men of age 50. The company estimates that the probability that a man of age 50 dies within a year is .01. Estimate the number of claims that the company can expect from beneficiaries of these men within a year.

$X \sim \text{Binomial}(n=1000, p=0.01)$

$E(X) = np = 1000 \times 0.01 = \underline{10}$

Therefore, the company can expect 10 claims within a year.

6.1

***31** (Feller[14]) A large number, N, of people are subjected to a blood test. This can be administered in two ways: (1) Each person can be tested separately, in this case N test are required, (2) the blood samples of k persons can be pooled and analyzed together. If this test is *negative,* this one test suffices for the k people. If the test is *positive,* each of the k persons must be tested separately, and in all, $k+1$ tests are required for the k people. Assume that the probability p that a test is positive is the same for all people and that these events are independent.

(a) Find the probability that the test for a pooled sample of k people will be positive.

(b) What is the expected value of the number X of tests necessary under plan (2)? (Assume that N is divisible by k.)

(c) For small p, show that the value of k which will minimize the expected number of tests under the second plan is approximately $1/\sqrt{p}$.

a) $P(1 \text{ person positive}) = p$

$P(1 \text{ person negative}) = 1-p$

$P(k \text{ people negative}) = (1-p)^k$ (people are independent)

$P(\text{pooled sample positive}) = 1 - P(\text{pooled sample negative})$

$= \underline{1-(1-p)^k}$

b) pooled test : ① negative $\to$ 1 test ② positive $\to$ k+1 test $\Big\}$ per group

N people, k people per group $\rightarrow \frac{N}{K}$ groups

$E(X) = \frac{N}{K} [1 \cdot P(\text{pooled sample negative}) + (k+1) \cdot P(\text{pooled sample positive})]$

$= \frac{N}{K} [(1-p)^k + (k+1) \cdot (1-(1-p)^k)]$

c) $E(X) = \frac{N}{K} [(1-p)^k + (k+1) \cdot (1-(1-p)^k)]$

After differentiating $E(X)$ on k,

set equation to zero when p is small,

to obtain the minimum.

6.1 ***38** (from Pittel[17]) Telephone books, n in number, are kept in a stack. The probability that the book numbered i (where $1 \leq i \leq n$) is consulted for a given phone call is $p_i > 0$, where the p_i's sum to 1. After a book is used, it is placed at the top of the stack. Assume that the calls are independent and evenly spaced, and that the system has been employed indefinitely far into the past. Let d_i be the average depth of book i in the stack. Show that $d_i \leq d_j$ whenever $p_i \geq p_j$. Thus, on the average, the more popular books have a tendency to be closer to the top of the stack. *Hint*: Let p_{ij} denote the probability that book i is above book j. Show that $p_{ij} = p_{ij}(1 - p_j) + p_{ji}p_i$.

Follow the solution and the hint:

P_{ij} is the probability that book i is above book j

$P_{ij} = \underbrace{P_{ij}(1-P_j)}_{①} + \underbrace{P_{ji} \cdot P_i}_{②}$

book i above book j has 2 conditions

① book i is already above j & book j is not select

② book j is above book i & book i is selected and now is above book j

$P_{ij} = P_{ij}(1-P_j) + (1-P_{ij}) \cdot P_i$

$P_{ji} = P(\text{book } j \text{ above book } i)$
$= 1 - P(\text{book } i \text{ above book } j)$
$= 1 - P_{ij}$

Then doing algebra,

$P_{ij} = P_{ij} - P_{ij} \cdot P_j + P_i - P_{ij} \cdot P_i$

$P_{ij} = P_{ij}(1 - P_j - P_i) + P_i$

$$1 = (1 - P_j - P_i) + \frac{P_i}{P_{ij}}$$

$$P_j + P_i = \frac{P_i}{P_{ij}}$$

$$P_{ij} = \frac{P_i}{P_i + P_j}$$

The average depth of book j is

$$d_j = \sum_{k \neq j} P_{kj}$$

$$d_j = \sum_{k \neq j} \frac{P_k}{P_k + P_j}$$

If $P_i \geq P_j \Rightarrow \frac{P_k}{P_k + P_i} \leq \frac{P_k}{P_k + P_j}$

$\Rightarrow d_i \leq d_j$

Another approach to this question is to list examples:

	B_1	B_2	B_3	B_4
S_1	P_{11}			P_{41}
S_2		P_{22}		
S_3			P_{33}	
S_4	P_{14}			P_{44}

e.g. $P_{11} = 0.4$, $P_{21} = 0.3$, $P_{31} = 0.2$, $P_{41} = 0.1$

$$P_{12} = 0.4\left(\frac{0.3}{1-0.3} + \frac{0.2}{1-0.2} + \frac{0.1}{1-0.1}\right)$$

then look for general patterns.

6.2

5 In a certain manufacturing process, the (Fahrenheit) temperature never varies by more than 2° from 62°. The temperature is, in fact, a random variable F with distribution

$$P_F = \begin{pmatrix} 60 & 61 & 62 & 63 & 64 \\ 1/10 & 2/10 & 4/10 & 2/10 & 1/10 \end{pmatrix}.$$ Discrete

(a) Find $E(F)$ and $V(F)$.

(b) Define $T = F - 62$. Find $E(T)$ and $V(T)$, and compare these answers with those in part (a).

(c) It is decided to report the temperature readings on a Celsius scale, that is, $C = (5/9)(F - 32)$. What is the expected value and variance for the readings now?

a) $E(F) = 60 \times \frac{1}{10} + 61 \times \frac{2}{10} + 62 \times \frac{4}{10} + 63 \times \frac{2}{10} + 64 \times \frac{1}{10}$

$= 6 + 12.2 + 24.8 + 12.6 + 6.4$

$= 62$

$E(F^2) = 60^2 \times \frac{1}{10} + 61^2 \times \frac{2}{10} + 62^2 \times \frac{4}{10} + 63^2 \times \frac{2}{10} + 64^2 \times \frac{1}{10}$

$= \frac{38452}{10}$

$V(F) = E(F^2) - [E(F)]^2$

$= \frac{38452}{10} - 62^2 = 1.2$

b) $E(T) = E(F-62) = E(F) - E(62) = 62 - 62 = 0$

$V(T) = V(F-62) = V(F) = 1.2$

c) $E(C) = E(\frac{5}{9}(F-32)) = E(\frac{5}{9}F - \frac{5}{9} \times 32)$

$= E(\frac{5}{9}F) - E(\frac{5}{9} \times 32)$

$= \frac{5}{9} E(F) - \frac{5}{9} \times 32$

$= \frac{5}{9} \times 62 - \frac{5}{9} \times 32 = 16.67$

$V(C) = V(\frac{5}{9}(F-32)) = V(\frac{5}{9}F)$

$= \frac{25}{81} V(F) = \frac{25}{81} \times 1.2 = 0.37$

6.3

7 Let X be a random variable with density function f_X. Show, using elementary calculus, that the function

$$\phi(a) = E((X-a)^2)$$

takes its minimum value when $a = \mu(X)$, and in that case $\phi(a) = \sigma^2(X)$.

$$f(a) = E((X-a)^2)$$
$$= \int (x-a)^2 f(x)\,dx$$

$$f'(a) = \int 2(x-a)(-1)f(x)\,dx$$
$$= \int (-2x+2a)f(x)\,dx$$
$$= -2\underbrace{\int x f(x)\,dx}_{E(x)} + 2a\underbrace{\int f(x)\,dx}_{=1}$$
$$= -2\mu(X) + 2a$$

$$f'(a) = 0$$
$$2\mu(x) = 2a$$
$$\underline{a = \mu(x)}$$