Markov and Cheybshev Inequalities and the Law of Large Numbers—Proofs and Illustrations

Markov's Inequality

Suppose a random variable X takes only nonnegative values so that $P\{X \geq 0\} = 1$.

How much probability is there in the tail of the distribution of X?

More specifically, for a given value $a > 0$, what can we say about the value of $P\{X \geq a\}$?

Markov's inequality takes $\mu = E(X)$ into account and provides an upper bound on $P\{X \geq a\}$ that depends on the values of a and μ.

We give the derivation for a continuous random variable X with density function $f(x)$:

$$\mu = E(X) = \int_{(0,\infty)} x\, f(x)\, dx = \int_{(0,a)} x\, f(x)\, dx + \int_{[\boldsymbol{a},\infty)} x\, f(x)\, dx$$

$$\geq \int_{[\boldsymbol{a},\infty)} x\, f(x)\, dx \geq \int_{[\boldsymbol{a},\infty)} a\, f(x) = a\int_{[\boldsymbol{a},\infty)} f(x)\, dx = aP\{X \geq a\},$$

from which we obtain Markov's Inequality:

$P\{X \geq a\} \leq \mu/a.$

In the above:

- The first inequality holds because
  the integral ignored is nonnegative,
- The second inequality holds because
  $a \le x$, for x in $[a, \infty)$.

The proof for a discrete random variable is similar,
with summations replacing integrals.

Markov's Inequality gives an
upper bound on $P\{X \ge a\}$ that applies to
any distribution with positive support.

Practical consequences.

For most distributions of practical interest,

the probability in the tail beyond a is

noticeably smaller than μ/a for all values of a.

Below, for several continuous and discrete distributions,
each with $\mu = 1$,
we use R to show that the nonincreasing "reliability function"
$R(a) = 1 - F(a) = P\{X > a\}$ is
bounded above by $\mu/a = 1/a$.

(For continuous distributions there is no difference between P$\{X \le a\}$ and P$\{X < a\}$, and for discrete distributions the discrepancy is not noticeable in our graphs.)

The Markov bound $1/a$ is not useful for $a < 1$ (that is $1/a > 1$) because no probability exceeds 1.

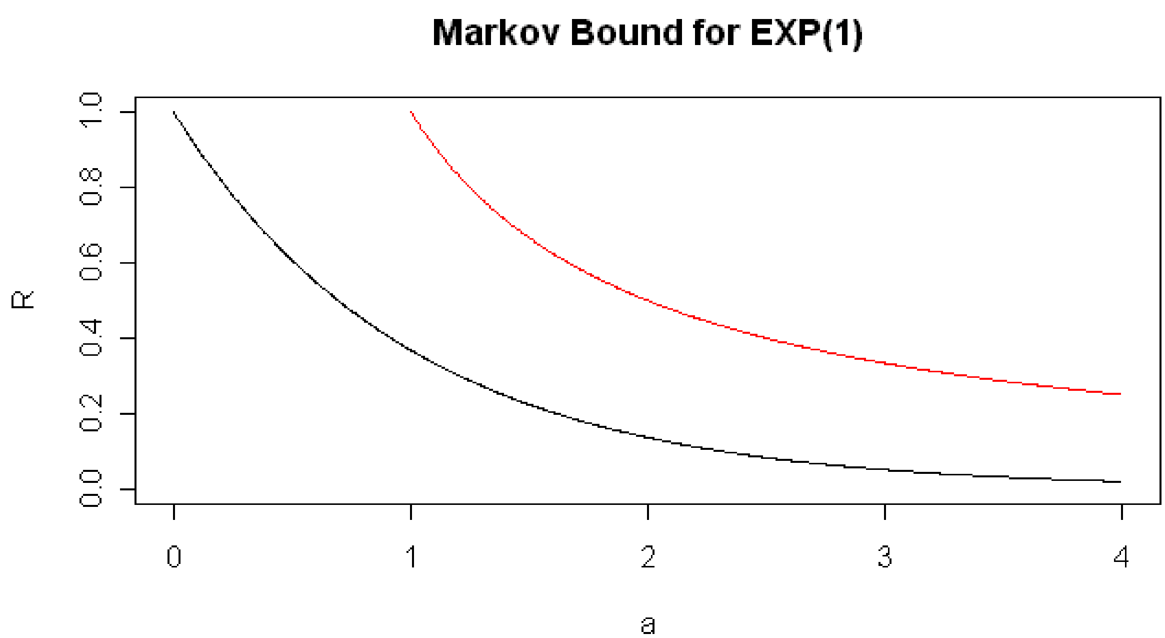
Problems:

(a) Verify the functional form of $R(a)$ in each example.

(b) Make a plot of R and the Markov bound for the degenerate random variable X with P$\{X = 1\} = 1$.

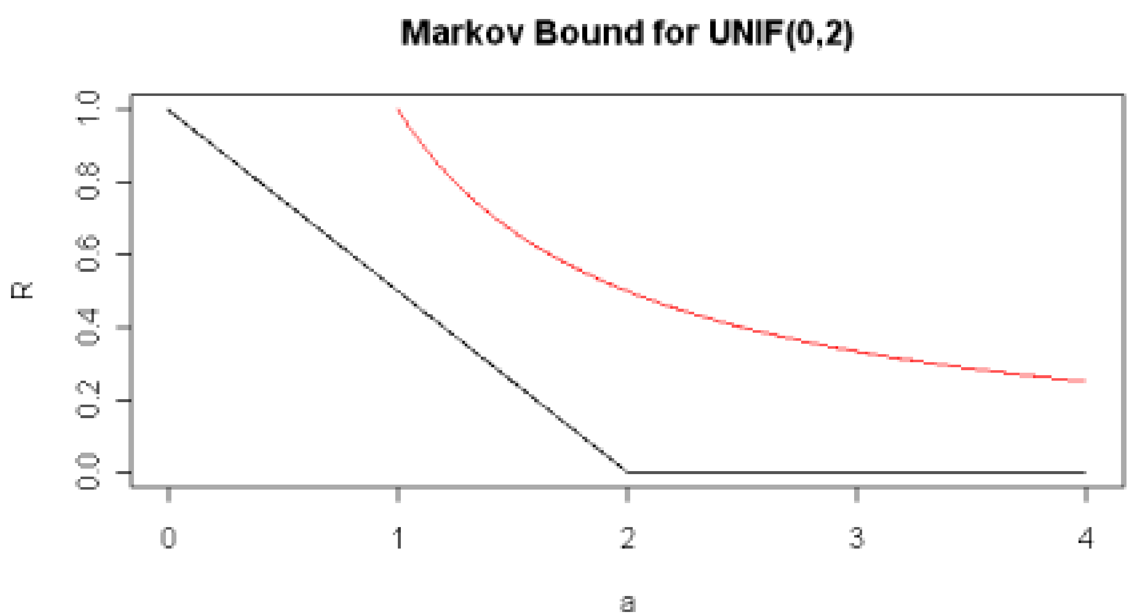
Exponential distribution with mean 1

```
a <- seq(0, 4, length=1000)
aa <- seq(1, 4, length= 500)
plot(a, 1 - pexp(a, rate=1), type="l",
    ylim=c(0,1), ylab="R",
    main="Markov Bound for EXP(1)")
lines(aa, 1/aa, col="red")
```



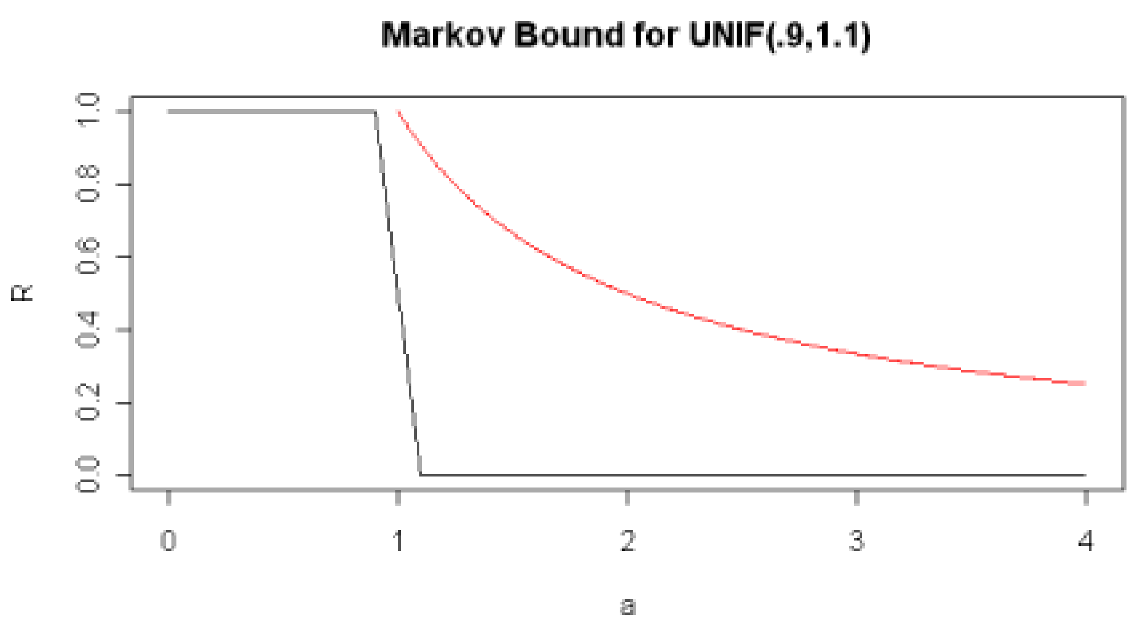
Uniform distribution on (0, 2).

```
a <- seq(0, 4, length=1000)
aa <- seq(1, 4, length= 500)
plot(a, 1 - punif(a,0,2), type="l",
    ylim=c(0,1), ylab="R",
    main="Markov Bound for UNIF(0,2)")
lines(aa, 1/aa, col="red")
```



Uniform distribution on (0.9,1.1).

```
a <- seq(0, 4, length=1000)
aa <- seq(1, 4, length= 500)
plot(a, 1 - punif(a,.9,1.1), type="l",
ylim=c(0,1), ylab="R", main="Markov Bound for
UNIF(.9,1.1)")
lines(aa, 1/aa, col="red")
```

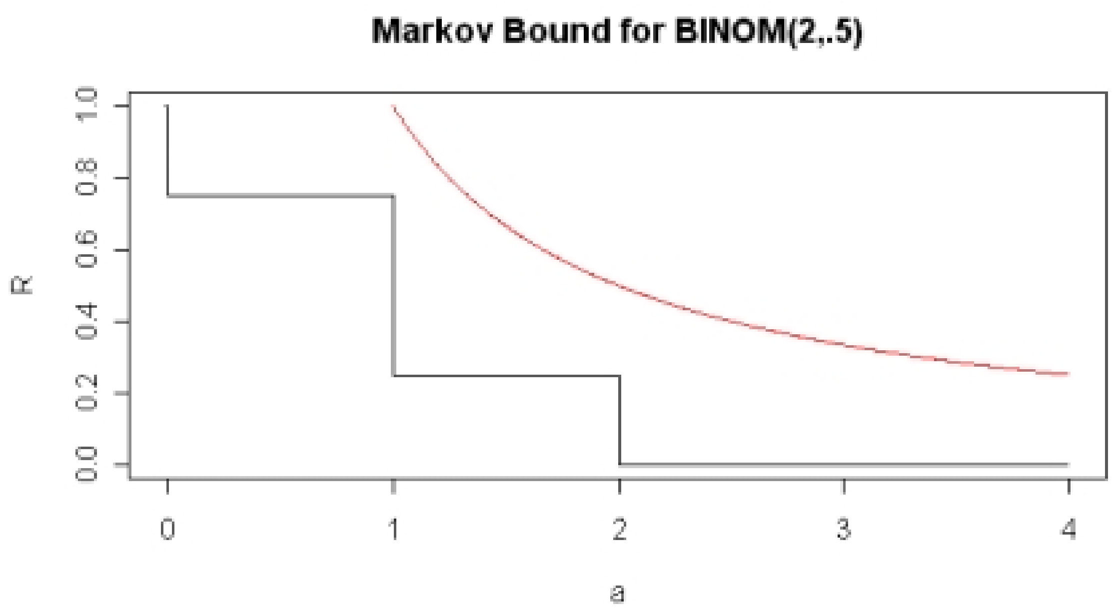


At a = 1, the bound nearly touches R. A distribution more tightly concentrated about μ = 1 would come even closer to touching.

So, as a general statement, the Markov bound cannot be improved.

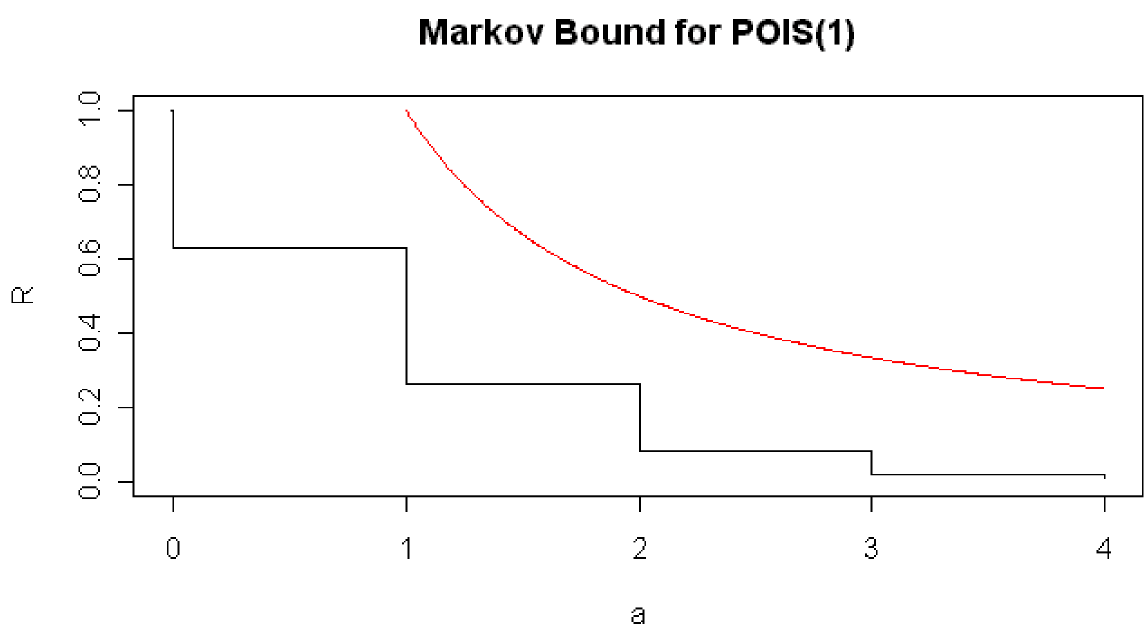
Binomial distribution with *n* = 2 and *p* = 1/2.

```
a <- seq(-.01, 4, by=.001)
aa <- seq(1, 4, length= 500)
plot(a, 1 - pbinom(a,2,.5), type="l",
    ylim=c(0,1), ylab="R",
    main="Markov Bound for BINOM(2,.5)")
lines(aa, 1/aa, col="red")
```



Poisson distribution with $\lambda = 1$.

```
a <- seq(-.01, 4, by=.001)
aa <- seq(1, 4, length= 500)
plot(a, 1 - ppois(a,1), type="l",
    ylim=c(0,1), ylab="R",
    main="Markov Bound for POIS(1)")
lines(aa, 1/aa, col="red")
```



Chebyshev's Inequality

Suppose a random variable Y has

$E(Y) = \mu$ and $V(Y) = \sigma^2 < \infty$.

Then, setting $X = (Y - \mu)^2$, we have

$\mu_X = E(X) = V(Y)$ and $P(X \geq 0) = 1$.

Now apply Markov's Inequality to X with $b > 0$ to obtain

$P\{X \geq b^2\} = P\{(Y - \mu)^2 \geq b^2\} = P\{|Y - \mu| \geq b\} \leq \sigma^2/b^2$.

Letting $b = k\sigma$, we obtain Chebyshev's Inequality:

$P\{|Y - \mu| \geq k\sigma\} \leq 1/k^2$,

which (by the complement rule) is sometimes written as

$P\{|Y - \mu| < k\sigma\} \geq 1 - 1/k^2$.

In words, not more than $1/k^2$ of the

probability in a distribution can lie

beyond k standard deviations

away from its mean.

The following table compares the some of the information from Chebyshev's Inequality with *exact information about a normal* distribution (upon which the Empirical Rule is based).

Interval	**Exact**	**Chebyshev**
$(\mu - \sigma, \mu + \sigma)$	**0.681**	**Uninformative**
$(\mu - 1.5\sigma, \mu + 1.5\sigma)$	**0.866**	$\geq 1 - 4/9 = 0.556$
$(\mu - 2\sigma, \mu + 2\sigma)$	**0.950**	$\geq 1 - 1/4 = 0.750$
$(\mu - 3\sigma, \mu + 3\sigma)$	**0.997**	$\geq 1 - 1/9 = 0.889$

Below for several continuous and discrete distributions,

each with $\mu = 1$ and $\sigma = 1$,

we use R to show that the nondecreasing function

$Q(k) = P\{|Y - \mu| \leq k\sigma\} = P\{|Y - 1| \leq k\}$ is

bounded above by $1 - 1/k^2$.

Problems:

(c) In the normal example below, verify the values in the table above as closely as you can by reading the graph.

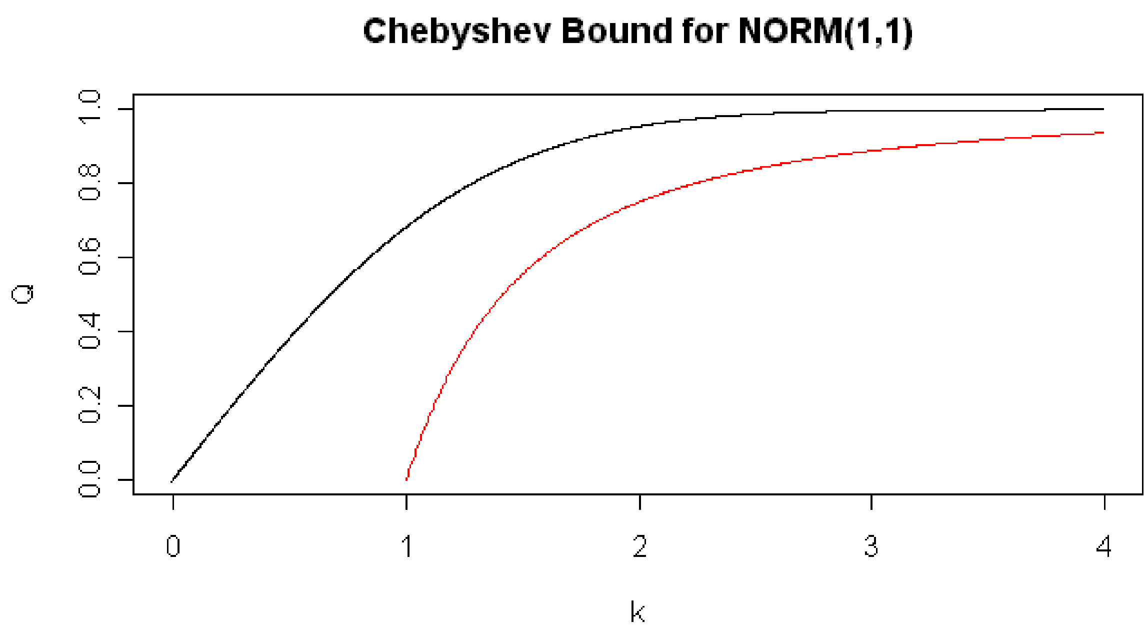
(d) In the remaining examples, verify that the variance is 1 and find the functional forms of Q. Draw a similar sketch of Q compared with the Chebyshev bound for the distribution that places probability 1/8 at each of the points 0 and 2 and probability 3/4 at 1.

Normal distribution with $\mu = 1$ and $\sigma = 1$.

```
k <- seq(-.01, 4, by=.001)

kk <- seq(1, 4, length= 500)

plot(k, pnorm(1+k,1,1)-pnorm(1-k,1,1),
  type="l", ylim=c(0,1), ylab="Q",
  main="Chebyshev Bound for NORM(1,1)")
```



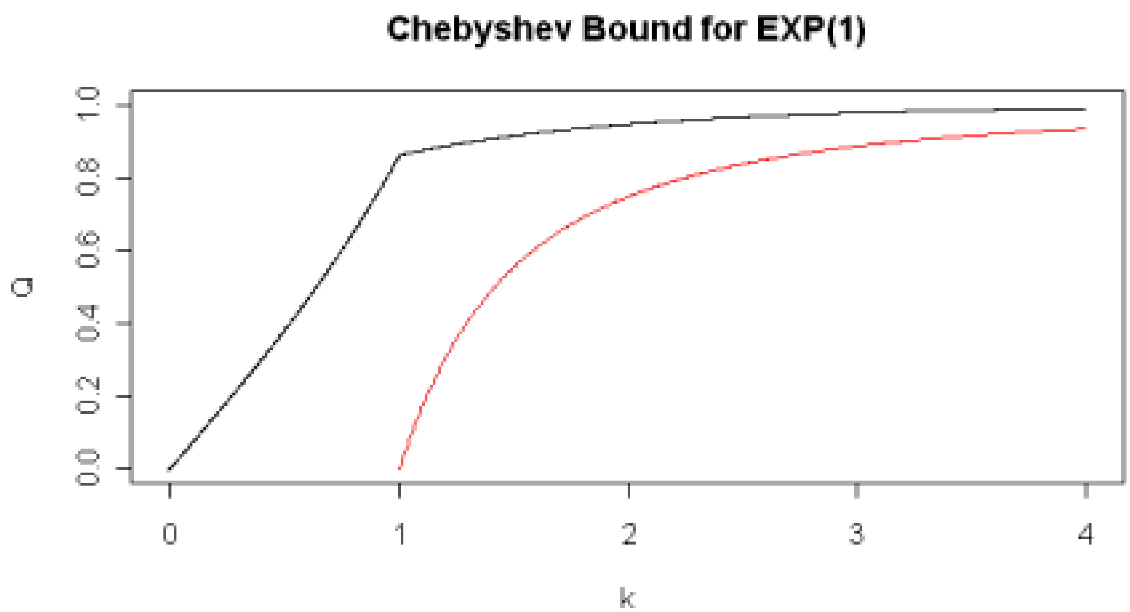

```
lines(kk, 1 - 1/kk^2, col="red")
```

Exponential distribution with mean 1.

```
k <- seq(-.01, 4, by=.001)

kk <- seq(1, 4, length= 500)

plot(k, pexp(1+k,1,1)-pexp(1-k,1,1), type="l",
  ylim=c(0,1), ylab="Q",
  main="Chebyshev Bound for EXP(1)")
```




```
lines(kk, 1 - 1/kk^2, col="red")
```

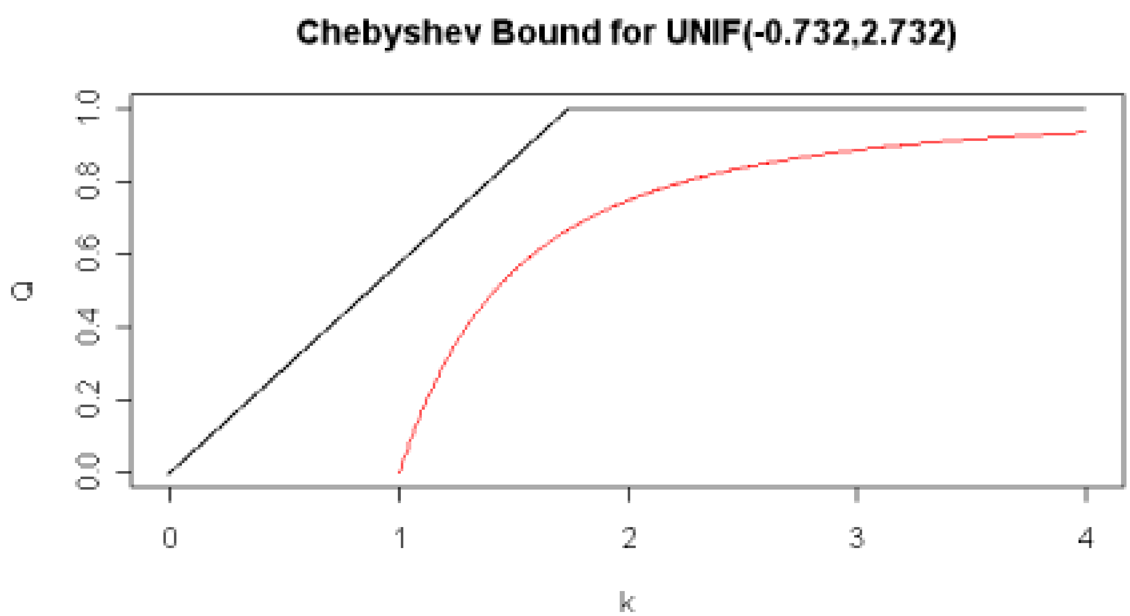
Uniform distribution on $(1 - \sqrt{3}, 1 + \sqrt{3})$.

```
k <- seq(-.01, 4, by=.001)

kk <- seq(1, 4, length= 500)

plot(k, punif(1+k,1-sqrt(3),1+sqrt(3))-punif(1-
  k,1-sqrt(3),1+sqrt(3)), type="l",
  ylim=c(0,1), ylab="Q", main="Chebyshev Bound
  for UNIF(-0.732,2.732)")

lines(kk, 1 - 1/kk^2, col="red")
```

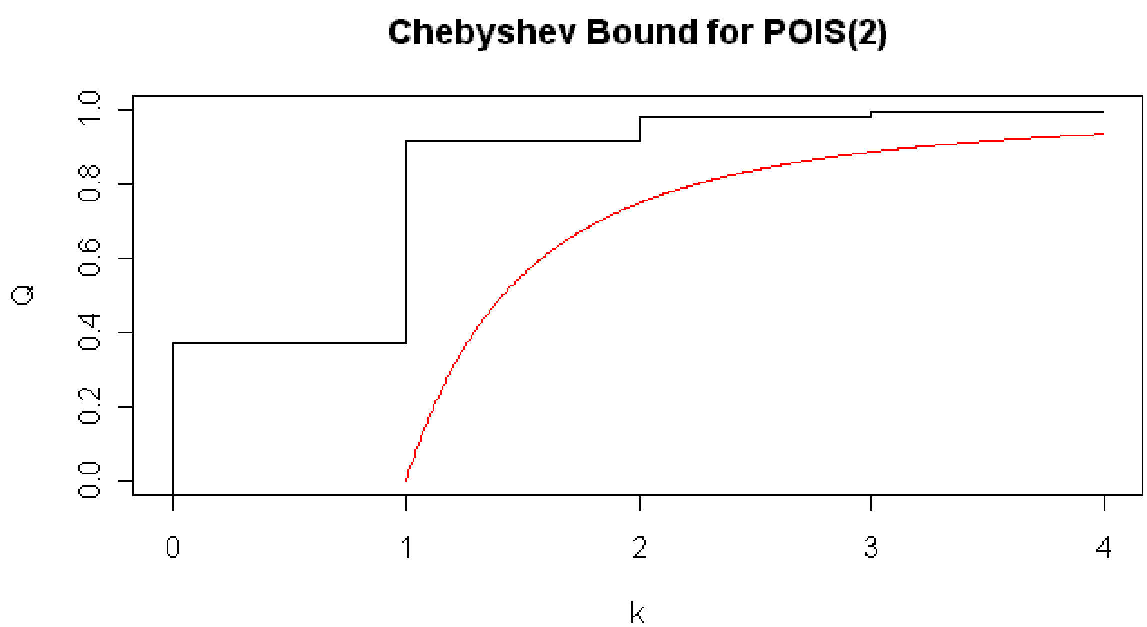


Poisson distribution with $\lambda = 1$.

```
k <- seq(-.01, 4, by=.001)

kk <- seq(1, 4, length= 500)

plot(k, ppois(1+k,1)-ppois(1-k-.001,1),
  type="l", ylim=c(0,1), ylab="Q",
  main="Chebyshev Bound for POIS(2)")
```




```
lines(kk, 1 - 1/kk^2, col="red")
```

For relatively large values of *k*, the Chebyshev bound can be used to get a reasonable, if rough, approximation to the probability in the tail(s) of a distribution.

Of course, when the exact distribution is known and its probability distribution is available as a function in R, it is preferable to get the exact value.

Law of Large Numbers for Coin Tossing.

An important theoretical use of Chebyshev's Inequality is to prove the Law of Large Numbers for Coin Tossing.

If a coin with P(Heads) = p is tossed n times, then the heads ratio Z_n = #(Heads)/n has mean $\mu = E(Z_n) = p$ and $\sigma^2 = V(Z_n) = p(1-p)/n$.

Thus for arbitrarily small $\varepsilon = k\sigma > 0$, Chebyshev's inequality gives

$$1 \geq P\{|Z_n - p| < \varepsilon\} \geq 1 - p(1-p)/n\varepsilon^2 \to 1, \qquad \text{as } n \to \infty.$$

Thus $P\{|Z_n - p| < \varepsilon\} \to 1$.

Note: Here $\varepsilon = k\sigma = k[p(1-p)/n]^{1/2}$, so $1/k^2 = p(1-p)/n\varepsilon^2$.

We say that Z_n converges to p in probability and write $Z_n \xrightarrow{prob} p$.

In words, as the number of tosses increases to a sufficiently large number, we see that the heads ratio is within ε of p with probability as near 1 as we please.

As a practical matter, Z_n is nearly normally distributed for large n. Thus the normal distribution is a better way than the Chebyshev bound to assess the accuracy of Z_n as an estimate of p.

Roughly speaking, this amounts to
using the Empirical Rule.

As a specific example:
in 10000 tosses of a fair coin,
$2\,\mathrm{SD}(V_{10000}) = 2(pq/10000)^{1/2} = 2(40000)^{-1/2} = 2/200 = 0.01.$

So we can expect the heads ratio to be
within 0.01 of 1/2
with probability 95%.

Fifty thousand tosses would allow
approximation of p with 2-place accuracy.

Problem:

(e) What does Chebyshev's inequality say about
$\mathrm{P}\{|Z_{12500} - p| < 1/150\}$?
What does the normal approximation say
about this probability?

The graph on the next page
illustrates Chebyshev and CLT bounds
for tossing a fair coin.

```
m <- 10000
alpha <- .01
p <- 1/2
q <- 1 - p
n <- 1:m
sd <- sqrt(p*q/n)
h <- rbinom(m,1, p)
s <- cumsum(h)
p.hat <- s/n
k.ch <- sqrt(1/alpha); k.clt <- qnorm(1 - alpha/2)
k.ch
k.clt
plot(n, p.hat, pch=".",
   ylim=c(max(0,p-.2), min(1,p+.2)),
   main="Coin Tosses with CLT (red) and Chebyshev Bounds")
lines(n, p+k.ch*sd, col="green")
lines(n, p-k.ch*sd, col="green")
lines(n, p+k.clt*sd, col="red")
lines(n, p-k.clt*sd, col="red")
```