Chapter 4: Probability inequalities and theorems

Markov's inequality

Suppose that $g(Y)$ is a function of random variable Y such that $\mathbb{E}[|g(Y)|] < \infty$. For any $k > 0$, it can be shown that

$$P(|g(Y)| \geq k) \leq \frac{\mathbb{E}[|g(Y)|]}{k}.$$

or equivalently

$$P(|g(Y)| < k) \geq 1 - \frac{\mathbb{E}[|g(Y)|]}{k}.$$

This is *Markov's inequality*.

Chebychev's inequality

aka

- Chebyshev's inequality
- Bienaymé-Chebyshev inequality
- Tchebysheff's Theorem
- Chebysheff, Chebychov, Chebyshov, Tchebychev . . .

Chebychev's inequality (cont.)

Suppose that Y has finite mean μ and variance σ^2.

In Markov's inequality, consider the function

$$g(t) = (t - \mu)^2.$$

We have for any $k > 0$ that

$$P(|(Y - \mu)^2| < k) \geq 1 - \frac{\mathbb{E}[|(Y - \mu)^2|]}{k}.$$

which can be more simply written as

$$P((Y - \mu)^2 < k) \geq 1 - \frac{\mathbb{E}[(Y - \mu)^2]}{k}.$$

Chebychev's inequality (cont.)

But

$$\mathbb{E}[(Y-\mu)^2] = \sigma^2$$

and if we choose to write $k = c^2\sigma^2$, the inequality becomes

$$P((Y-\mu)^2 < c^2\sigma^2) \geq 1 - \frac{\mathbb{E}[(Y-\mu)^2]}{c^2\sigma^2} = 1 - \frac{1}{c^2}$$

Chebychev's inequality (cont.)

The left hand side of the inequality is equal to

$$P(\sqrt{(Y-\mu)^2} < \sqrt{c^2\sigma^2}) = P(|Y-\mu| < c\sigma)$$

so the inequality becomes

$$P(|Y-\mu| < c\sigma) \geq 1 - \frac{1}{c^2}.$$

or equivalently

$$P(|Y-\mu| \geq c\sigma) \leq \frac{1}{c^2}.$$

This is *Chebychev's inequality*.

Chebychev's inequality (cont.)

This states that for *any* distribution with finite mean μ and variance σ^2, we have that

$$P(\mu - c\sigma < Y < \mu + c\sigma) \geq 1 - \frac{1}{c^2}$$

which bounds how much probability can be present in the tails of the distribution.

The sample mean

Suppose that $Y_1, \ldots, Y_n$ are independent that have the same distribution, with expectation μ and variance σ^2. If

$$U = \sum_{i=1}^{n} Y_i$$

then

$$\mathbb{E}[U] = n\mu \qquad \mathbb{V}[U] = n\sigma^2.$$

Let

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

be the *sample mean* random variable.

The sample mean (cont.)

Then

$$\mathbb{E}[\overline{Y}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \frac{1}{n} n\mu = \mu$$

and it can be shown that

$$\mathbb{V}[\overline{Y}_n] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

which follows by *independence*.

The sample mean (cont.)

Now recall Chebychev's inequality:

$$P(|Y - \mu| \geq c\sigma) \leq \frac{1}{c^2}.$$

Applying this to $\overline{Y}_n$, we have that

$$P\left(|\overline{Y}_n - \mu| \geq \frac{c\sigma}{\sqrt{n}}\right) \leq \frac{1}{c^2}.$$

The sample mean (cont.)

We can rewrite this for any n by setting

$$\epsilon = \frac{c\sigma}{\sqrt{n}} \quad \text{so that} \quad c = \frac{\sqrt{n}\epsilon}{\sigma}.$$

Then the inequality becomes

$$P(|\overline{Y}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Convergence in probability

Consider a sequence of random variables $Y_1, Y_2, \ldots, Y_n, \ldots$, and real constant c. We say that the sequence *converges in probability* to c as $n \longrightarrow \infty$ if, for every $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P(|Y_n - c| < \epsilon) = 1$$

that is, the probability

$$P(c - \epsilon < Y_n < c + \epsilon)$$

converges to 1 as $n \longrightarrow \infty$. We write

$$Y_n \xrightarrow{p} c$$

as $n \longrightarrow \infty$ to indicate convergence in probability.

Convergence in probability (cont.)

Note

- If, for a fixed $\epsilon > 0$,

$$p_n = P(|Y_n - c| < \epsilon)$$

the definition requires that the real sequence $\{p_n\}$ converges to 1 in the usual sense.

- There is no restriction on the joint distribution of the Y_is in the sequence (that is, they can be dependent).
- We can equivalently write the condition

$$\lim_{n \longrightarrow \infty} P(|Y_n - c| > \epsilon) = 0.$$

The weak law of large numbers

Suppose that $Y_1, Y_2, \ldots, Y_n,$ are independent with the same distribution, and expectation μ and variance σ^2.

Then

$$\overline{Y}_n \xrightarrow{p} \mu$$

as $n \longrightarrow \infty$.

The weak law of large numbers (cont.)

This follows from Chebychev's inequality: recall that for the sample mean

$$P(|\overline{Y}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

or equivalently

$$P(|\overline{Y}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}.$$

for any $\epsilon > 0$.

The weak law of large numbers (cont.)

But

$$1 - \frac{\sigma^2}{n\epsilon^2} \longrightarrow 1$$

as $n \longrightarrow \infty$, as σ and ϵ are fixed. Thus

$$\lim_{n \longrightarrow \infty} P(|\overline{Y}_n - \mu| < \epsilon) \geq 1$$

and hence $\overline{Y}_n \xrightarrow{p} \mu$.

The weak law of large numbers (cont.)

What this result means: under these conditions, the distribution of sample mean random variable becomes *degenerate* at expectation of the distribution of the Ys.

- For example, if you repeatedly and independently toss a fair coin, the proportion of heads will converge to 1/2 as the number of tosses increases.

Convergence in distribution

For cdfs, $F_1(y), F_2(y), \ldots, F_n(y), \ldots$, suppose that

$$\lim_{n \longrightarrow \infty} F_n(y) = F(y)$$

for some cdf $F(y)$ at all points where $F(y)$ is *continuous*.

Then $F(y)$ is the *limiting distribution* for the sequence.

Convergence in distribution (cont.)

If $Y_1, Y_2, \ldots, Y_n, \ldots$ are a sequence of random variables having cdfs $F_1(y), F_2(y), \ldots, F_n(y), \ldots$ respectively, and Y has cdf $F(y)$, then we say that the sequence $Y_1, Y_2, \ldots, Y_n, \ldots$ *converges in distribution* to Y,

$$Y_n \xrightarrow{d} Y$$

as $n \longrightarrow \infty$.

The central limit theorem

Suppose that $Y_1, Y_2, \ldots, Y_n$, are independent, have the same distribution, with finite expectation μ and variance σ^2.

For each n, let

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

be the sample mean random variable.

The central limit theorem (cont.)

Then the random variable

$$U_n = \frac{\sqrt{n}(\overline{Y}_n - \mu)}{\sigma}$$

converges in distribution to a *standard normal distribution*.

That is,

$$\lim_{n \longrightarrow \infty} F_{U_n}(u) = \lim_{n \longrightarrow \infty} P(U_n \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

This result is the *Central Limit Theorem* (CLT).

The central limit theorem (cont.)

Note that we could write

$$U_n = \frac{(\overline{Y}_n - \mu)}{\sigma/\sqrt{n}}$$

or, multiplying numerator and denominator by n

$$U_n = \frac{\left(\sum_{i=1}^{n} Y_i - n\mu\right)}{\sqrt{n}\sigma}$$

The central limit theorem (cont.)

The result says that, provided the conditions on expectation and variance are met, when n is large, we can approximate the distribution of $\overline{Y}_n$, or of

$$S_n = \sum_{i=1}^{n} Y_i$$

using a Normal distribution, irrespective of the actual distribution of $Y_1, \ldots, Y_n$.

The central limit theorem (cont.)

That is,

- $\overline{Y}_n$ is approximately distributed as

$$Normal(\mu, \sigma^2/n)$$

- S_n is approximately distributed as

$$Normal(n\mu, n\sigma^2)$$

The central limit theorem (cont.)

Note

If $Y_1, \ldots, Y_n$ are independent and identically distributed Normal random variables with expectation μ and variance σ^2, then

- $\overline{Y}_n$ is exactly distributed as

$$Normal(\mu, \sigma^2/n)$$

- S_n is exactly distributed as

$$Normal(n\mu, n\sigma^2)$$

Applications of the central limit theorem

We can use the central limit theorem to approximate some standard distributions for which it is otherwise awkward to compute probabilities.

Suppose we wish to approximate the distribution of random variable Y. We must first attempt to decompose Y as

$$Y = \sum_{i=1}^{n} Y_i$$

where $Y_1, \ldots, Y_n$ are independent and identically distributed, and have expectation μ and variance σ^2 that can be easily computed.

Applications of the central limit theorem (cont.)

We then can say that Y is approximately distributed as

$$Normal(n\mu, n\sigma^2).$$

and compute, for example

$$F_Y(y) = \Pr(Y \leq y) = P\left(\frac{Y - n\mu}{\sqrt{n}\sigma} \leq \frac{y - n\mu}{\sqrt{n}\sigma}\right) \simeq \Phi\left(\frac{y - n\mu}{\sqrt{n}\sigma}\right)$$

where $\Phi(.)$ is the standard Normal cdf.

Applications of the central limit theorem (cont.)

Example (Binomial case)

If $Y \sim Binomial(n, p)$, the mgf of Y takes the form

$$m_Y(t) = (pe^t + q)^n$$

By the uniqueness of mgfs, we recognize that Y has the representation

$$Y = \sum_{i=1}^{n} Y_i$$

where $Y_i \sim Binomial(1, p) \equiv Bernoulli(p)$ are independent. We have

$$\mathbb{E}[Y_i] = p \qquad \mathbb{V}[Y_i] = pq.$$

Applications of the central limit theorem (cont.)

Example (Binomial case)

Therefore for large n, Y is approximately distributed as

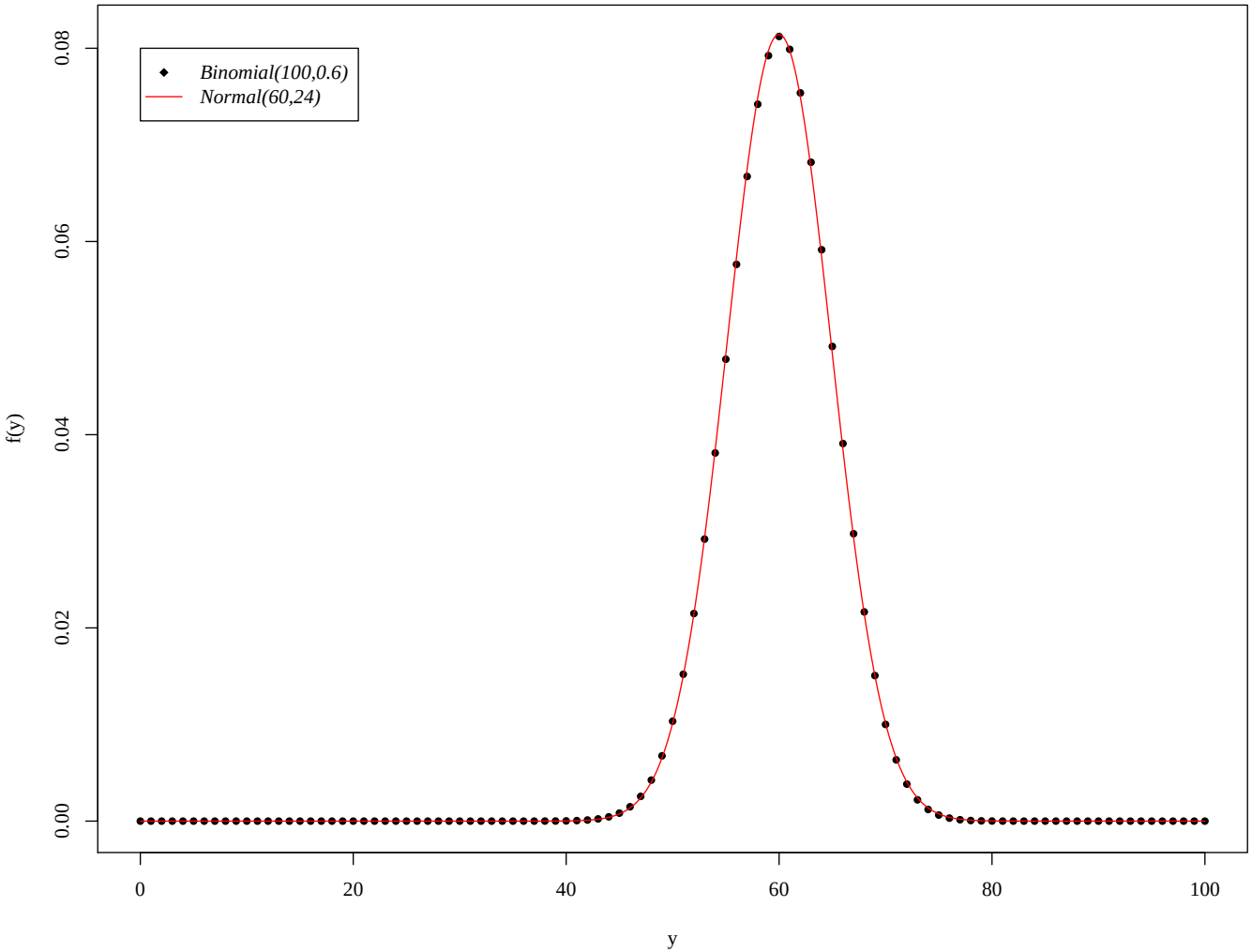
$$Normal(np, npq)$$

and

$$F_Y(y) = P(Y \leq y) \simeq \Phi\left(\frac{y - np}{\sqrt{npq}}\right)$$

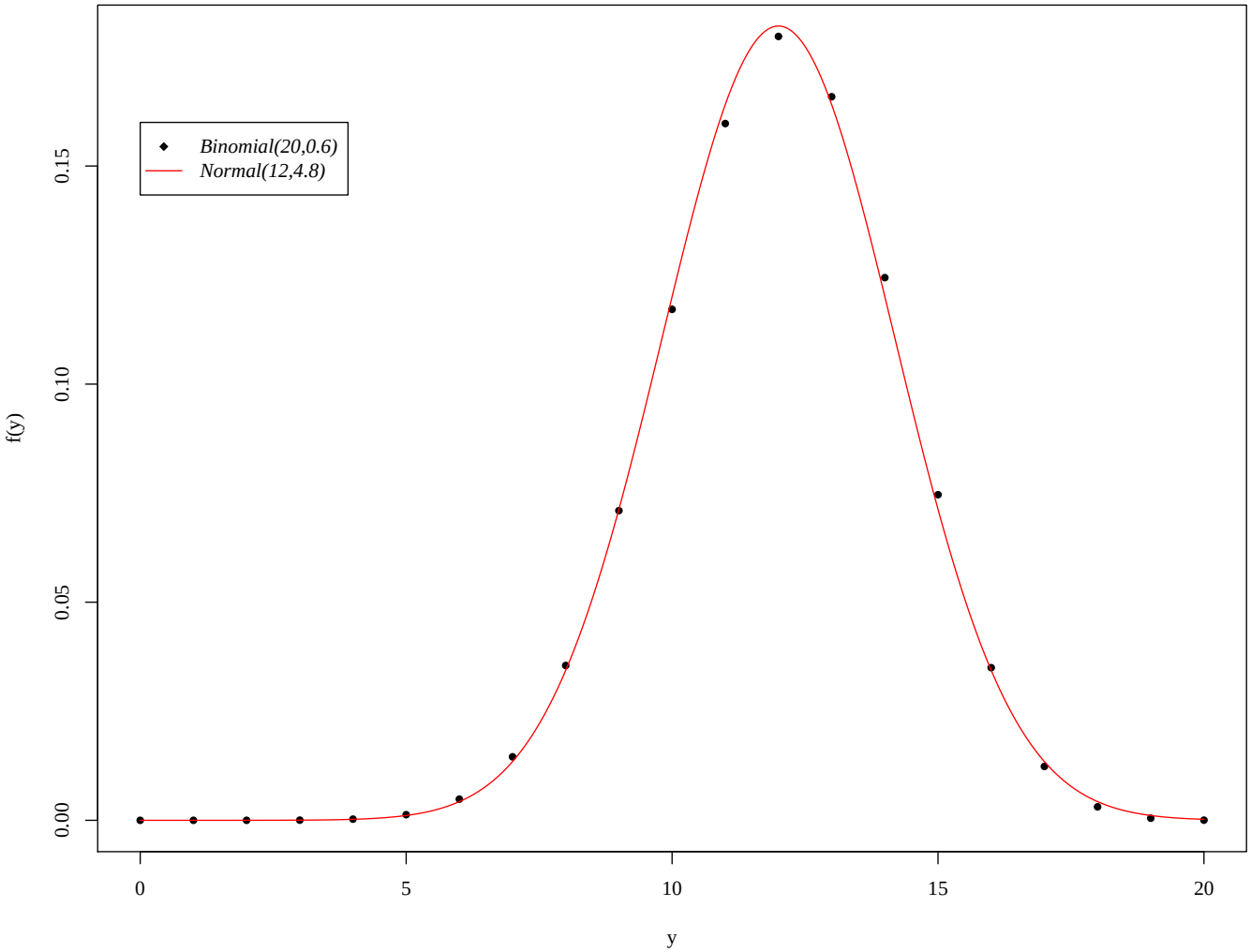
Applications of the central limit theorem (cont.)

Binomial: $n = 100$, with $p = 0.6$.



Applications of the central limit theorem (cont.)

Binomial: $n = 20$, with $p = 0.6$.



Applications of the central limit theorem (cont.)

Example (Poisson case)

If $Y \sim Poisson(\lambda)$, the mgf of Y takes the form

$$m_Y(t) = \exp\{\lambda(e^t - 1)\}.$$

By the uniqueness of mgfs, we recognize that Y has the representation

$$Y = \sum_{i=1}^{n} Y_i$$

where the $Y_i \sim Poisson(\lambda/n)$ are independent. We have

$$\mathbb{E}[Y_i] = \frac{\lambda}{n} \qquad \mathbb{V}[Y_i] = \frac{\lambda}{n}.$$

Applications of the central limit theorem (cont.)

Example (Poisson case)

Therefore Y is approximately distributed as

$$Normal(\lambda, \lambda)$$

and

$$F_Y(y) = P(Y \leq y) \simeq \Phi\left(\frac{y-\lambda}{\sqrt{\lambda}}\right)$$

Note that for this approximation to work, we need to have λ large.

Applications of the central limit theorem (cont.)

Continuity correction

When we use the central limit theorem to approximate a discrete distribution, a *continuity correction* is sometimes applied: that is, if Y is discrete we compute

$$F_Y(y) = P(Y \leq y) \simeq \Phi\left(\frac{y + \frac{1}{2} - n\mu}{\sqrt{n}\sigma}\right)$$

Applications of the central limit theorem (cont.)

Example (Gamma case)

If $Y \sim Gamma(\alpha, \beta)$, the mgf of Y takes the form

$$m_Y(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha}$$

By the uniqueness of mgfs, we recognize that Y has the representation

$$Y = \sum_{i=1}^{n} Y_i$$

where the $Y_i \sim Gamma(\alpha/n, \beta)$ are independent. We have

$$\mathbb{E}[Y_i] = \frac{\alpha\beta}{n} \qquad \mathbb{V}[Y_i] = \frac{\alpha\beta^2}{n}.$$

Applications of the central limit theorem (cont.)

Example (Gamma case)

Therefore Y is approximately distributed as

$$Normal(\alpha\beta, \alpha\beta^2)$$

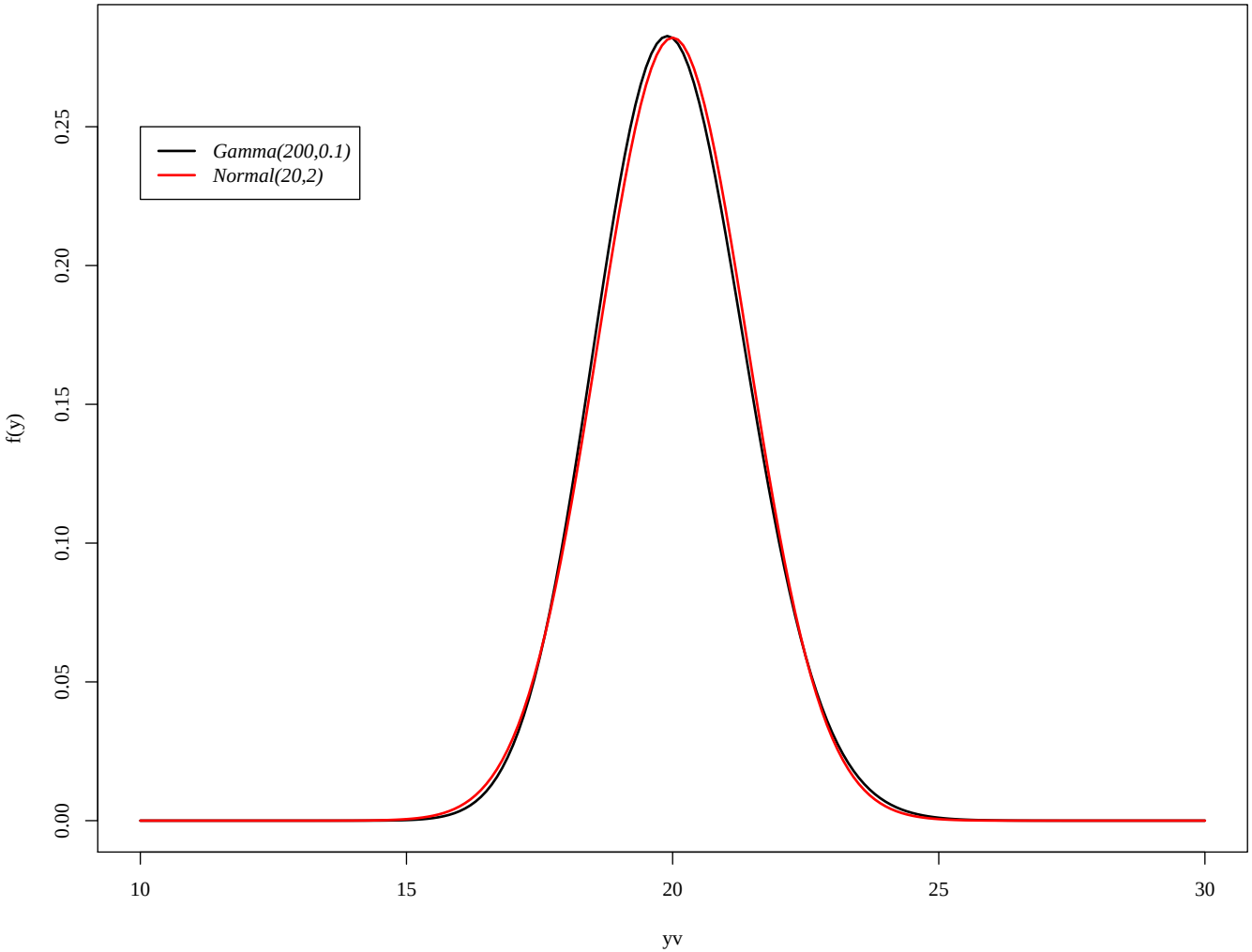
and

$$F_Y(y) = P(Y \leq y) \doteq \Phi\left(\frac{y - \alpha\beta}{\sqrt{\alpha\beta^2}}\right)$$

Note that for this approximation to work, we need to have α large.

Applications of the central limit theorem (cont.)

Gamma: $\alpha = 200$, with $\beta = 0.1$.



Applications of the central limit theorem (cont.)

Gamma: $\alpha = 50$, with $\beta = 0.4$.

