Normal Distribution

Paul E. Johnson <pauljohn@ku.edu>

January 26, 2006

1   Mathematical Description

The Normal distribution describes a continuous variable that takes on values in the real number line. The formula for the Normal has two **parameters**, μ and σ^2. The parameter μ is a "location" parameter and σ^2 is a "scale" parameter. The probability density function is often written as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

I think it looks a little nicer if rearranged.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

If you want to make sense out of that, it is vital for you to have a mental image of the function e^{-z^2} on the interval between 0 and 5:
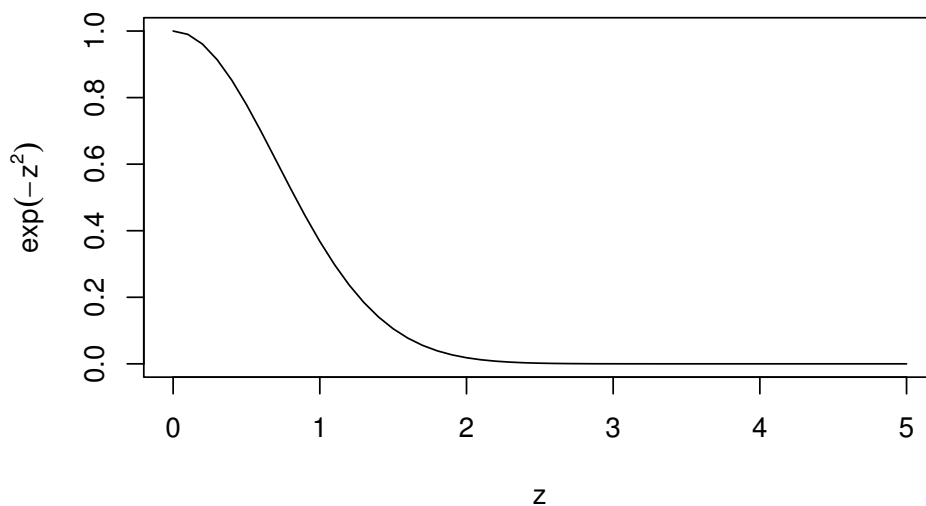
Note well that "all of the action" in the Normal probability formula is in the e^{-z^2} part, not in the "normalizing constant" $\frac{1}{\sqrt{2\pi\sigma^2}}$.

In maximum likelihood, we usually end up trying to maximize the log of the likelihood function, and when the Normal is logged, then it simplifies quite dramatically:

$$
\begin{aligned}
ln(p(x)) &= ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] + ln\left[e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right] \\
&= \left[ln(1) - ln(\sqrt{2\pi\sigma^2})\right] - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \\
&\quad 0 - ln((2\pi\sigma^2)^{1/2}) - \frac{2}{2\sigma^2}(x-\mu)^2 \\
&= -\frac{1}{2}ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2 \\
&\quad -\frac{1}{2}ln(2) - \frac{1}{2}ln(\pi) - ln(\sigma) - \frac{1}{2\sigma^2}(x-\mu)^2
\end{aligned}
$$

2   Illustrations

The probability density function of a Normal distribution with $\mu = 3$ and $\sigma = 5$ is shown in Figure 1. The R code which produces that figure is:

```
> mu <- 3
> sigma <- 5
> xrange <- seq(from = mu - 3 * sigma, to = mu + 3 * sigma, by = 0.2)
> mainlabel <- expression(paste("Normal Distribution, N(", mu,
+      ",", sigma, "^2", ")", sep = ""))
> xprob <- dnorm(xrange, mean = mu, sd = sigma, log = F)
> plot(xrange, xprob, type = "l", main = mainlabel, xlab = "possible values of x",
+      ylab = "probability of x")
```

How would one describe that? Well, off the top of my head, I'd say the most outstanding characteristics are that it is:

1. Unimodal

2. Symmetric

How does this distribution change in appearance if μ and σ^2 are changed? Let's do some experimentation. The following R code creates an array of figures with 4 rows and 2 columns with various values of μ and σ.
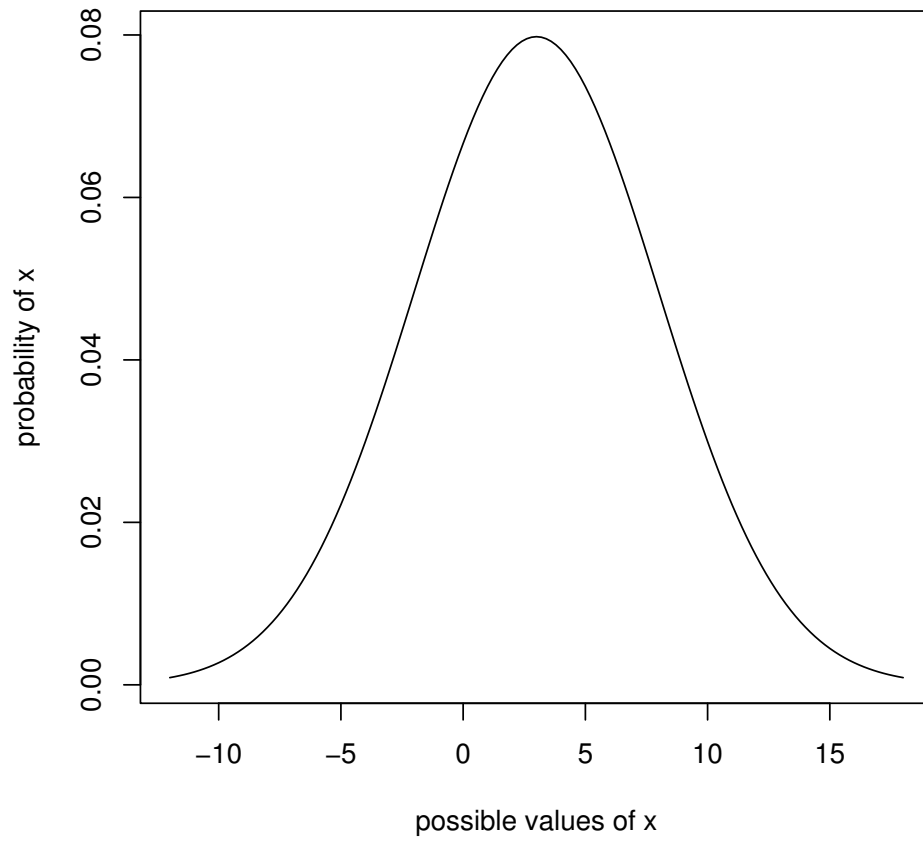
```
> par(mfrow = c(4, 2))
> for (i in 1:4) {
+      for (j in 1:2) {
+          mu = 3 * j
```

Normal Distribution, N(μ,σ^2)

```
+            sigma <- 2 * i
+            xrange <- seq(from = mu - 3 * sigma, to = mu + 3 * sigma,
+                by = 0.2)
+            mainlabel <- paste("N(", mu, ",", sigma * sigma, ",)",
+                sep = "")
+            xprob <- dnorm(xrange, mean = mu, sd = sigma, log = F)
+            plot(xrange, xprob, type = "l", main = mainlabel, xlab = "possible values of
+                ylab = "probability of x")
+        }
+ }
```

2

While Figure 2 is quite boring and repetitious, it does convey one very important attribute of the Normal distribution: **it always keeps the same shape**. At least for these parameter values, it is unimodal and symmetric. These graphs look the same because the X axis is allowed to re-scale itself to use up the allocated space.

If we restrict the display so that the axes of all of the figures are kept the same–in a position that suits the largest set of values–then the impact of changing the parameters is a bit more apparent. The code only needs to be modified very slightly by a specification of the xlim option in the plot statement.

```
> par(mfrow = c(4, 2))
> for (i in 1:4) {
+     for (j in 1:2) {
+         mu = 3 * j
+         sigma <- 2 * i
+         xrange <- seq(from = mu - 3 * sigma, to = mu + 3 * sigma,
+             by = 0.2)
+         mainlabel <- paste("N(", mu, ",", sigma * sigma, ",)",
+             sep = "")
+         xprob <- dnorm(xrange, mean = mu, sd = sigma, log = F)
+         plot(xrange, xprob, type = "l", main = mainlabel, xlab = "possible values of
+             ylab = "probability of x", xlim = c(-20, 30))
+     }
+ }
```

The result is to be seen in Figure 3.

3   What about the parameters μ and σ^2?

As already mentioned, the parameter μ indicates the "location", the left-right position of the distribution. The parameter σ is a "scale" parameter, determining how far it reaches from left to right.

Most students are already aware of the fact that μ and σ actually play a more familiar role. In fact, many students tell me that the mean of anything is represented by a parameter μ and standard deviation is σ. But it is not always so.

Figure 2: Variety of Normals

N(3,4,)

probability of x

possible values of x

N(6,4,)

probability of x

possible values of x

N(3,16,)

probability of x

possible values of x

N(6,16,)

probability of x

possible values of x

N(3,36,)

probability of x

possible values of x

N(6,36,)

probability of x

possible values of x

N(3,64,)

probability of x

possible values of x

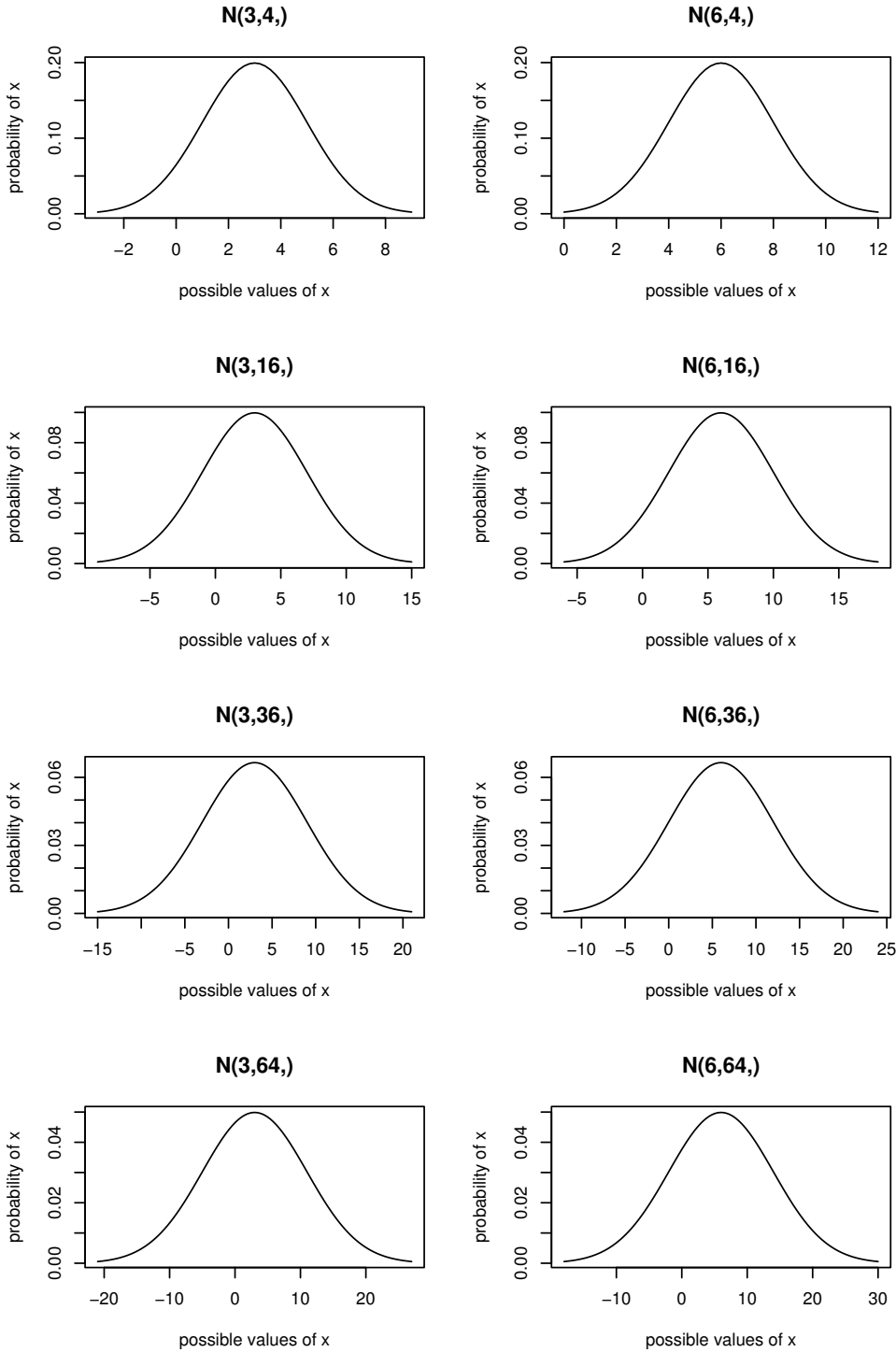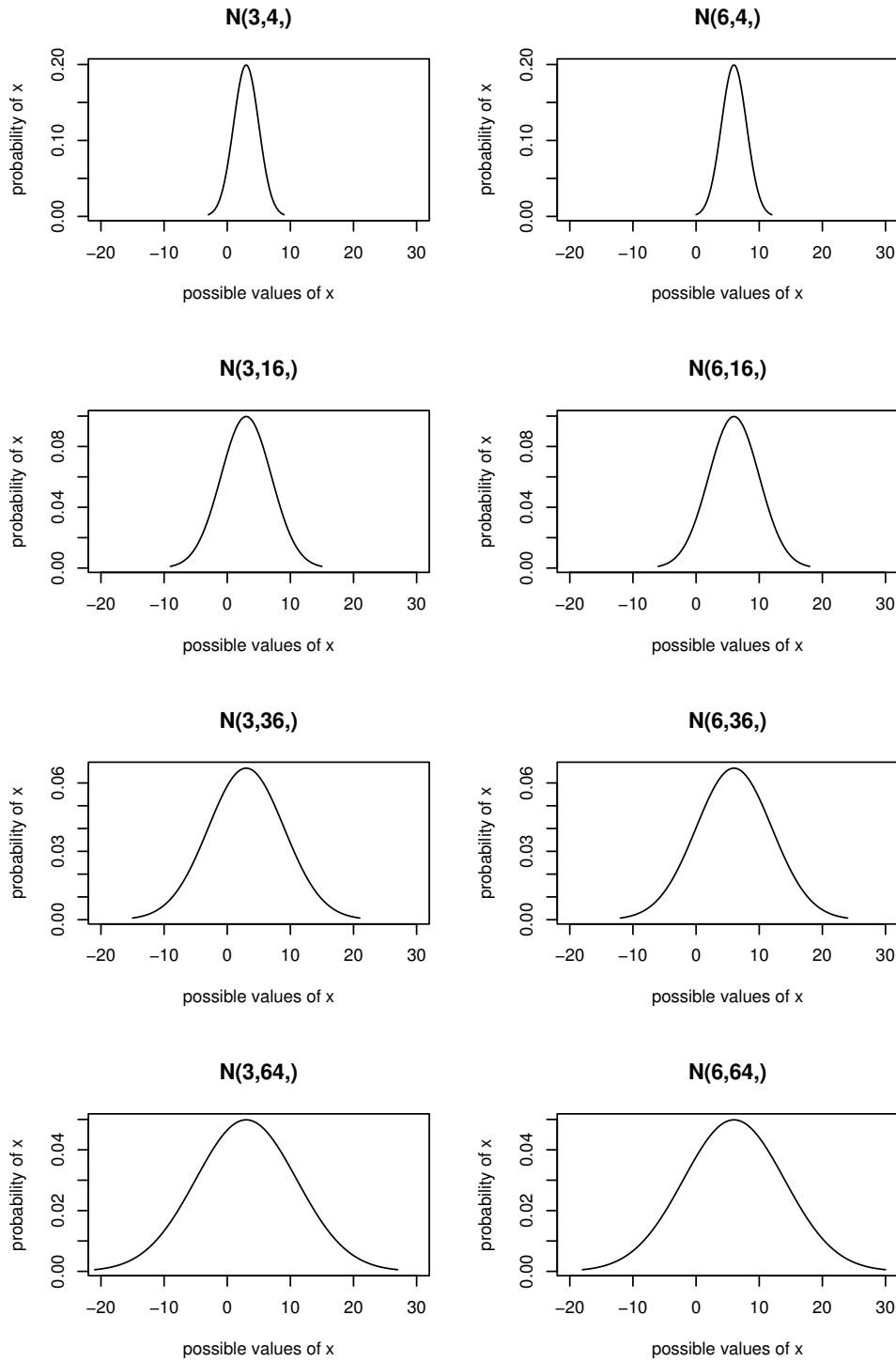**N(6,64,)**

probability of x

possible values of x

Figure 3: Normals (again)

Now, in case you are not one of those students, here's some background. The **expected value** of a distribution is defined as the "probability weighted sum" of outcomes. For $x \sim N(\mu, \sigma^2)$,

$$E(x) = \int_{-\infty}^{+\infty} p(x) \cdot x \, dx$$

and, through the magic of mathematics, it turns out that

$$E(x) = \mu$$

How do you find that? Well, honestly, I think most of us just look up the answer in a book! If you really want to calculate it, get a book on mathematical statistics and find out how to use a "moment generating function" to calculate the expected values of distributions.

The **variance** of a distribution is the "probability weighted sum" of the squared differences between outcomes and their expected values.

$$Var(x) = \int_{-\infty}^{+\infty} p(x) \cdot [x - E(x)]^2 \, dx$$

which can be rearranged as

$$Var(x) = \int_{-\infty}^{+\infty} p(x) \cdot x^2 \, dx - E(x)^2 = E(x^2) - E(x)^2$$

Repeat out loud: "The Variance of x equals the Expected value of x squared minus the Expected value of x, squared."

Again through the magic of mathematics, one would find

$$Var(x) = \sigma^2$$

I think my point here is that it is a completely fortuitous thing that the parameters μ and σ^2 are simply equal to the expected value and variance (respectively). This is quite rare in the gamut of statistical distributions. The expected value and variance are almost always equal to some more complicated formulae which combine the parameters.