Chapter 4: Models for the Random Error

Introductory Statistics for Engineering Experimentation

Peter R. Nelson, Marie Coffin and Karen A.F. Copeland

Slides by Douglas Bates

Outline

Outline

Outline

Outline

Random variables

- Models for the random error are called *random variables*. These are functions that map experimental outcomes to numeric values. Although we don't know what the particular value will be, we can describe the set of all possible values and characterize the distribution of the random variable.
- Distributions are characterized by either a *probability density function* (pdf) or a *probability mass function* (pmf) depending on whether the random variable is *continuous* (can take on any value in an interval) or *discrete* (its possible values are distinct - usually integers representing counts).
- The random variable is written in upper case. We will use a script font, like $\mathcal{Y}$. A particular value is written in lower case, like y.

Density functions for continuous random variables

- For a continuous random variable, $\mathcal{Y}$, the probability density function, written $f(y)$, is a non-negative function whose *integral* is unity. That is

$$f(y) \geq 0, \quad \text{all } y$$

$$\int_{-\infty}^{\infty} f(y)\, dy = 1$$

The probability that $\mathcal{Y}$ falls in an interval is the integral of the density over the interval

$$P[a \leq \mathcal{Y} \leq b] = \int_{a}^{b} f(y)\, dy$$

Probability mass functions for discrete random variables

- A discrete random variable has positive probability for only a finite or "countably infinite" set of values. We call this set "the set of all possible values" of the random variable.
- The probability mass function, sometimes called more simply "the probability function" is $P(\mathcal{Y} = k) = p_k$. We require

$$p_k \geq 0, \quad \text{all } k$$

$$\sum_{\text{all possible } k} p_k = 1$$

Mean and Variance of a Random Variable

- Just as we define the mean (average) and variance of a sample we have similar concepts for a random variable.
- The mean, written μ, is a weighted average using the probability function or the probability density to define the weights. (For a density an "average" is calculated by integrating.)
- For a discrete random variable (r.v.), the mean, or "expected value" is

$$\mu = \mathrm{E}(\mathcal{Y}) = \sum_{\text{all possible } k} k\, P(\mathcal{Y} = k) = \sum_{k} k\, p_k$$

and the variance, or mean squared deviation, is

$$\mathrm{Var}(\mathcal{Y}) = \sum_{k} (k - \mathrm{E}(\mathcal{Y}))^2\, P(\mathcal{Y} = k)$$

Mean and variance of continuous r.v.'s

- For a continous r.v. the mean and variance are

$$\mathrm{E}(\mathcal{Y}) = \int_{-\infty}^{\infty} y\, f(y)\, dy$$

$$\mathrm{Var}(\mathcal{Y}) = \int_{-\infty}^{\infty} (y - \mathrm{E}(\mathcal{Y}))^2\, f(y)\, dy$$

- Example 4.1.2 is based on a continuous uniform distribution on the interval $[0, 4]$ with $f(y) = 0.25, 0 \leq y \leq 4$ and 0 otherwise. Then

$$\mathrm{E}(\mathcal{Y}) = \int_{-\infty}^{\infty} y\, f(y)\, dy = \int_0^4 \frac{y}{4}\, dy = \left.\frac{y^2}{8}\right|_0^4 = 2$$

- Just as the sample mean is the point at which the sample can be balanced, the mean of a continuous r.v. is the point at which the density can be balanced. For a symmetric density like this it is the point of symmetry.

Properties of expected values

- In general we define the expected value of a function of a random variable, say $g(\mathcal{Y})$, written $\mathrm{E}\left(g(\mathcal{Y})\right)$, as either $\int_{-\infty}^{\infty} g(y)f(y)\,dy$ or $\sum_k g(k)p_k$, as appropriate.
- The calculation of $\mathrm{E}(\mathcal{Y})$ and $\mathrm{Var}(\mathcal{Y})$ can get complicated. Sometimes we can make things simpler by using rules related to "linear combinations". For example

$$\mathrm{E}(a + b\mathcal{Y}) = a + b\mathrm{E}(\mathcal{Y})$$

and when we consider two random variables, say $\mathcal{X}$ and $\mathcal{Y}$, then

$$\mathrm{E}(a\mathcal{X} + b\mathcal{Y}) = a\mathrm{E}(\mathcal{X}) + b\mathrm{E}(\mathcal{Y})$$

These results follow from the corresponding properties of sums or integrals.

Properties of expected values and variances

- The expected value of a constant function, say c, written $\mathrm{E}(c)$ is always c, no matter what the random variable.
- Two random variables are said to be *independent* if the outcome of one does not affect the outcome of the other.
- The variance of a constant, c, is always 0 (because it is the expected value of $(c-c)^2$).
- When you multiply a random variable by c you multiply its variance by c^2 (recall that the variance is on the scale of the square of the response). That is,

$$\mathrm{Var}(c\mathcal{Y}) = c^2\mathrm{Var}(\mathcal{Y})$$

- For independent random variables $\mathcal{X}$ and $\mathcal{Y}$,

$$\mathrm{Var}(\mathcal{X}+\mathcal{Y}) = \mathrm{Var}(\mathcal{X}) + \mathrm{Var}(\mathcal{Y})$$

Things to watch for

- Note that variances add, even when you subtract the random variables. For independent $\mathcal{X}$ and $\mathcal{Y}$

$$\mathrm{Var}(\mathcal{X} - \mathcal{Y}) = \mathrm{Var}(\mathcal{X} + (-1)\mathcal{Y}) = \mathrm{Var}(\mathcal{X}) + \mathrm{Var}(\mathcal{Y})$$

- Because a variance is an expected value of a square, which must be ≥ 0,

$$\mathrm{Var}(\mathcal{Y}) \geq 0, \quad \text{for any } \mathcal{Y}$$

- If you evaluate a variance and it comes out negative you have done something wrong.

Determining probabilities

- Probabilities can be considered as the long-term relative frequency of the outcomes of the experiment, assuming that it can be performed over and over again.
- In general we will refer to an experimental outcome or a set of outcomes as an *event*. We use upper-case Latin letters, like A, to denote events. We must have

$$0 \leq P(A) \leq 1.$$

- For discrete random variables, we can evaluate $P(A)$ as the sum of the probabilities of individual outcomes, p_k, for all $k \in A$. This uses the property that the probability of *mutually exclusive* (without overlap) events is the sum of the probabilities.
- One simple model for probabilities when there are a finite number of outcomes is *equally likely outcomes*. That is, each of the M outcomes has probability $\frac{1}{M}$.

Equally likely outcomes

- A favorite example of equally-likely outcomes is rolling a die. There are 6 possible outcomes, $1, \ldots, 6$ and, for a fair die, they are equally likely.
- Using this, if A is the event of getting an odd number, then

$$P(A) = p_1 + p_3 + p_5 = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

- Note that not all experiments with a finite set of outcomes have equally likely outcomes. In fact, very few do.
- If you roll two dice the possible totals are $2, \ldots, 12$ but they are not equally likely, even for fair dice. In this case $p_2 = \frac{1}{36}$ but $p_7 = \frac{1}{6}$.

Evaluating probabilities for continuous r.v.'s

- Recall that for a continuous random variable, $\mathcal{Y}$,

$$P(a \leq \mathcal{Y} \leq b) = \int_a^b f(y)\, dy$$

- Because the probability is expressed as an integral and the value of the function at a single point doesn't change the value of the integral, we have

$$P(a \leq \mathcal{Y} \leq b) = P(a \leq \mathcal{Y} < b) = P(a < \mathcal{Y} \leq b) = \ldots$$

Distribution functions

- The *cumulative distribution function* (cdf), $F(y)$, of a random variable, $\mathcal{Y}$, is

$$F(y) = P(\mathcal{Y} \leq y)$$

  Sometimes it is called just "the distribution function".
- The cdf is defined for discrete and for continuous random variables.
- It is always true that

$$P(a < \mathcal{Y} \leq b) = F(b) - F(a)$$

- For continuous random variables you can use $\leq$ in place of $<$ and vice versa and it is still true (see previous slide).
- For discrete random variables you can't interchange $\leq$ and $<$. You must be careful of which end points are in the interval.

The Bernoulli Distribution

- The Bernoulli distribution is the simplest, non-trivial probability distribution. There are two possible outcomes in the experiment which we arbitrarily label "success" and "failure". $\mathcal{Y}$ is the number of successes in 1 trial and must be either 0 or 1. We write $P(\mathcal{Y} = 1) = p$ which implies that $P(\mathcal{Y} = 0) = 1 - p$.
- The value $p \in [0, 1]$ is the *parameter* of this distribution. (This is one of the few cases where we use a Latin letter, instead of a Greek letter, for a parameter. Some texts use π for this parameter but that ends up being even more confusing than using p.)
- We can easily establish that $\mathrm{E}(\mathcal{Y}) = p$ and $\mathrm{Var}(\mathcal{Y}) = p(1 - p)$
- This distribution is mostly used as a building block for other distributions.

The Binomial Distribution

- If $\mathcal{Y}$ is the total number of successes in n independent, identical Bernoulli trials then we say that $\mathcal{Y}$ has a binomial distribution with

$$P(\mathcal{Y} = k) = p_k = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, \ldots, n$$

(Note that 0 is a possible value of $\mathcal{Y}$.)

- The symbol $\binom{n}{k}$ denotes the *binomial coefficient* which has the value

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- The *R* function `choose` evaluates binomial coefficients.

```
> choose(4, 2)
```

```
[1] 6
```

R functions for the binomial

- There are four *R* functions associated with the binomial distribution: `dbinom` evaluates the probability function, `pbinom` evaluates the cdf, `qbinom` evaluates the quantile function (the inverse of the cdf, sort-of) and `rbinom` returns a random sample from a binomial distribution.
- Most questions on the distribution itself involve evaluating the probability of some event (subset of the possible values of $\mathcal{Y}$). I prefer to do this as `sum(dbinom(range, size = n, prob = p))`
- In example 4.2.2, we have a binomial with $n = 50$ and $p = 0.03$ and we want $P(\mathcal{Y} < 2)$. This is

```
> sum(dbinom(0:1, 50, 0.03))
[1] 0.5552799
```

Properties of the binomial distribution

- By considering the binomial as the sum of n independent Bernoulli trials, each with probability p of success, we can derive

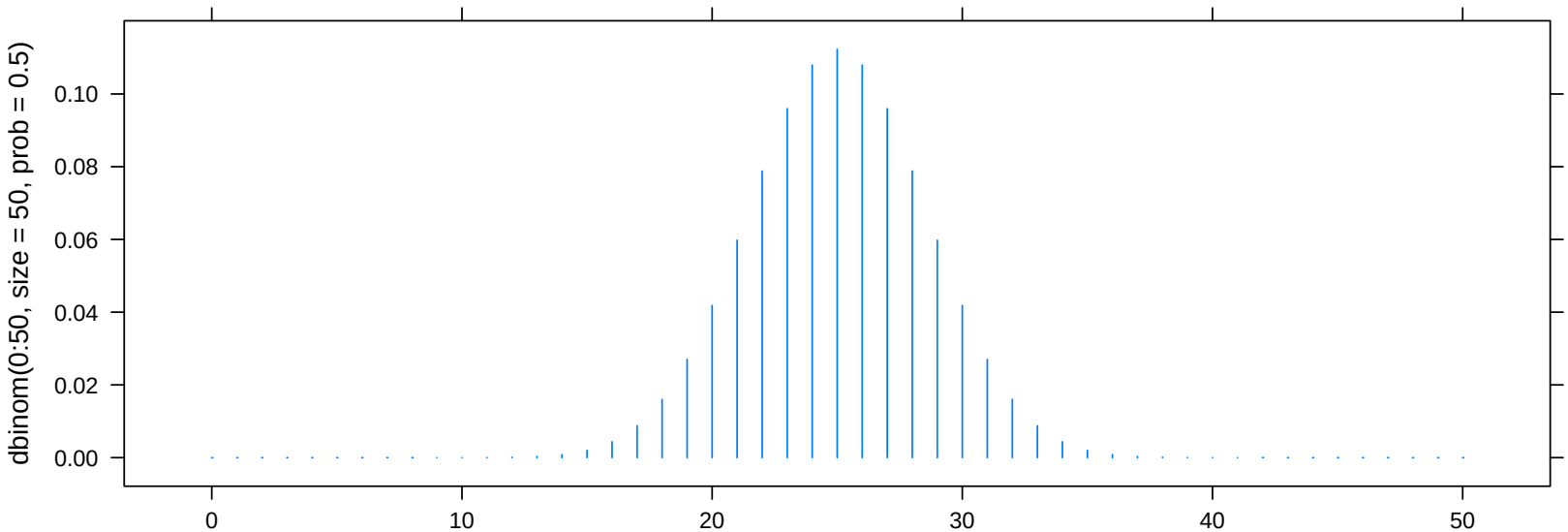
$$\mathrm{E}(\mathcal{Y}) = np$$
$$\mathrm{Var}(\mathcal{Y}) = np(1-p)$$

- The formula np for the expected value should make intuitive sense. It says that the expected number of successes is the number of trials times the probability of success on each trial.
- Note that the formula for the variance is symmetric in p and $1-p$. This has to be the case because you can change the meaning of "success" and "failure" without changing the variability in the distribution.
- When p is close to 0.5, the probability function is symmetric. When p is close to 0 or 1 the probability function is skew.
- For a fixed number of trials, n, the variance is greatest when $p = 0.5$. As p approaches 0 or 1 the variance goes to zero.

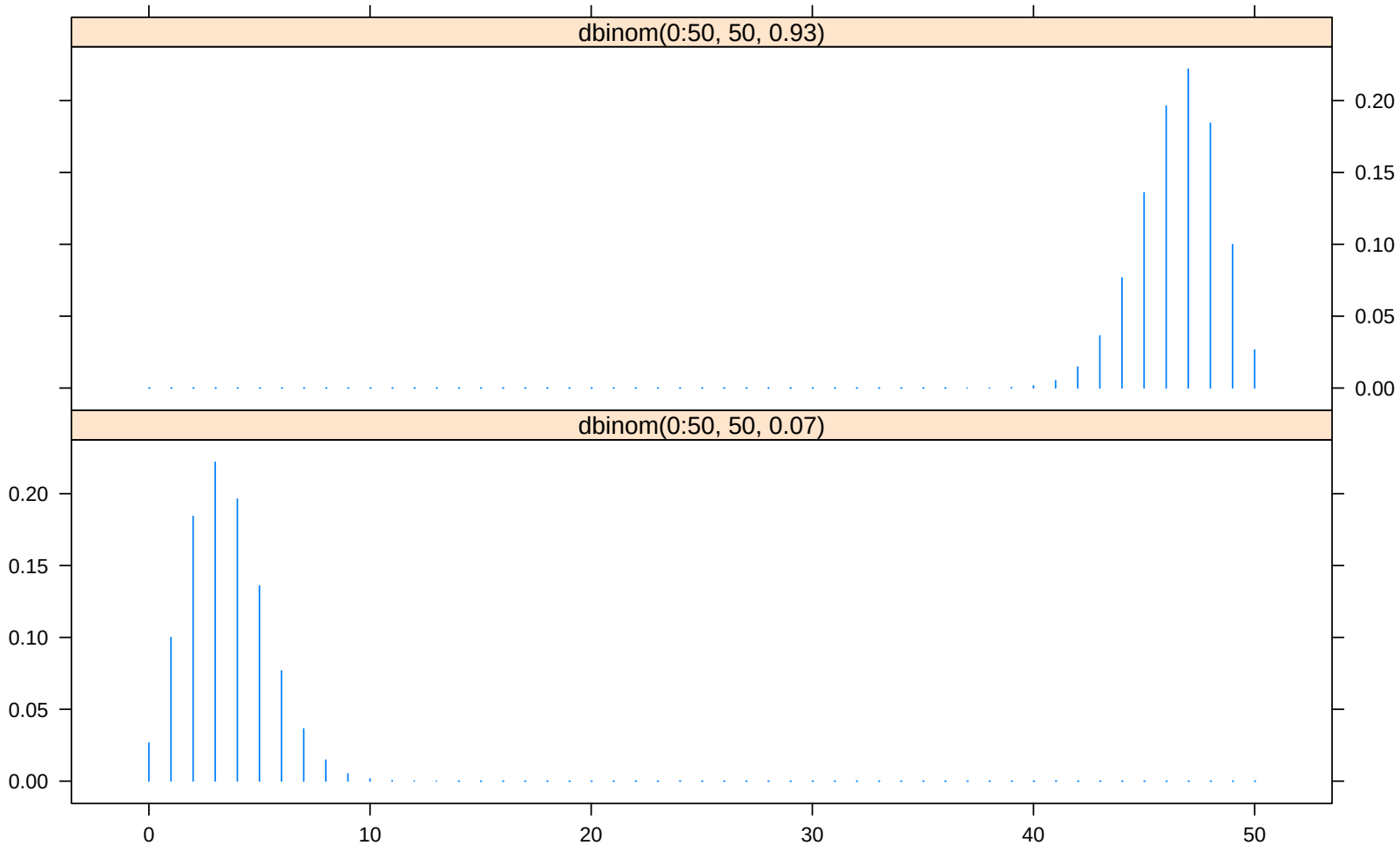
Plotting discrete probability functions

- One value for the `type` parameter in `xyplot` is `"h"` for "high-density" lines. This is the usual way of plotting a probability function for a discrete random variable.

```
> xyplot(dbinom(0:50, size = 50, prob = 0.5) ~ 0:50, type = "h")
```



Interchanging "success" and "failure"



Parameter estimation

- When we have a probability model for some observed data we formulate *estimates* for the parameters using the data values.
- The estimates are *statistics*, which describes any value that you can calculate based on the data alone. The formula for the estimate is called the *estimator*.
- For a binomial, the data are y, the number of successes, and n, the number of trials. The parameter is p, the probability of success on each trial. Not surprisingly, our "best guess" or point estimate for p is the observed proportion of successes.

$$\hat{p} = \frac{y}{n}$$

The Poisson distribution

- Like the binomial distribution, the Poisson distribution models the total number of events that occur in some experiment.
- For the binomial distribution the events are binary responses in distinct trials. The range of the random variable is $[0, n]$ where n is the total number of trials.
- In a *Poisson process* events occur over a continuum, such as time or distance or area or .... We assume that
  1. The interval of interest can be divided into small units of opportunity h such that only one event could occur during a unit.
  2. For an opportunity unit h the probability of one event is λh. The parameter λ is called the intensity.
  3. The occurrence or non-occurence of events in non-overlapping intervals are independent.

The Poisson distribution (cont'd)

- Let $\mathcal{Y}$ be the number of events in an opportunity unit of size t in a Poisson process with intensity λ. Then

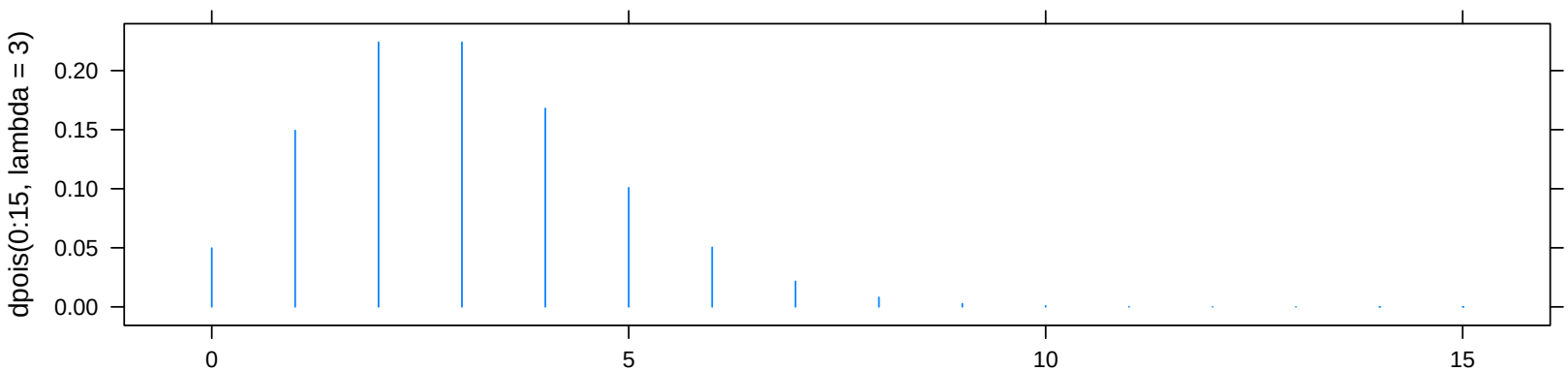
$$P(\mathcal{Y} = k) = p_k = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \quad k = 1, \ldots$$

- Most commonly t is a length of time (say 1 hr.) so that λ is the intensity of the events (say 30 cars/hr.). The parameter has units that are counts/(units of t).
- The *R* functions `dpois`, `ppois`, `qpois` and `rpois` take a single parameter, also called `lambda`, but representing the product λt in the notation above. This is the expected value of $\mathcal{Y}$.

$$\mathrm{E}(\mathcal{Y}) = \lambda t \quad \mathrm{Var}(\mathcal{Y}) = \lambda t$$

Example 4.2.7

- Example 4.2.6 describes a Poisson process where $\lambda = 0.1$ cars/sec. Suppose we observe for $t = 30$ seconds. Then $\lambda t = 3$ is the expected number of cars to observe.
- We are asked to evaluate the "probability that no more than 2 cars pass through the intersection", which is $P(\mathcal{Y} \leq 2)$.
- The simple evaluation is `sum(dpois(0:2, lambda = 3))` = 0.42319. An alternative is to use the cumulative distribution function `ppois(2, lambda = 3)` = 0.42319.



Parameter estimation

- Because $\mathrm{E}(\mathcal{Y}) = \lambda t$ we estimate the intensity, λ, by plugging into this relationship. That is, $\hat{\lambda} t = y$ or $\hat{\lambda} = y/t$.
- An interesting property of the Poisson distribution is the fact that the sum of independent Poisson random variables is also a Poisson. (Surprisingly this doesn't happen for most other distributions - only for the Poisson and the Gaussian or "normal" distributions.)
- Because

$$\mathcal{Y}_1 \sim \mathrm{Pois}(\lambda_1 t_1),\ \mathcal{Y}_2 \sim \mathrm{Pois}(\lambda_2 t_2) \Rightarrow \mathcal{Y}_1 + \mathcal{Y}_2 \sim \mathrm{Pois}(\lambda_1 t_1 + \lambda_2 t_2)$$

we can combine the results from non-overlapping intervals when estimating a common intensity λ.

Relationship to binomial

- The Poisson process can be considered as the continuous limit of discrete Bernoulli trials. The usual limit processes from calculus apply: take more and more intervals whose width becomes smaller and smaller to go from the discrete jumps to the continuous function.
- When binomial probabilities needed to be evaluated from tables, it was important that for n large and p moderate (so that both np and $n(1-p)$ become large) you could approximate binomial distributions by Poissons. We set $\lambda t = np$.
- We don't need to concern ourselves with that because we can evaluate essentially any binomial probability.

Example of Poisson approximation to binomial

- Just for completeness we give the comparison of $P(\mathcal{Y} < 2)$ for $n = 50$ and $p = 0.03$ versus a Poisson with $\lambda t = 1.5$.

```
> c(binom = sum(dbinom(0:1, 50, 0.03)), Poisson = sum(dpois(0:1,

    binom   Poisson
0.5552799 0.5578254

> all.equal(sum(dbinom(0:1, 50, 0.03)), sum(dpois(0:1, 1.5)))

[1] "Mean relative difference: 0.004584224"
```

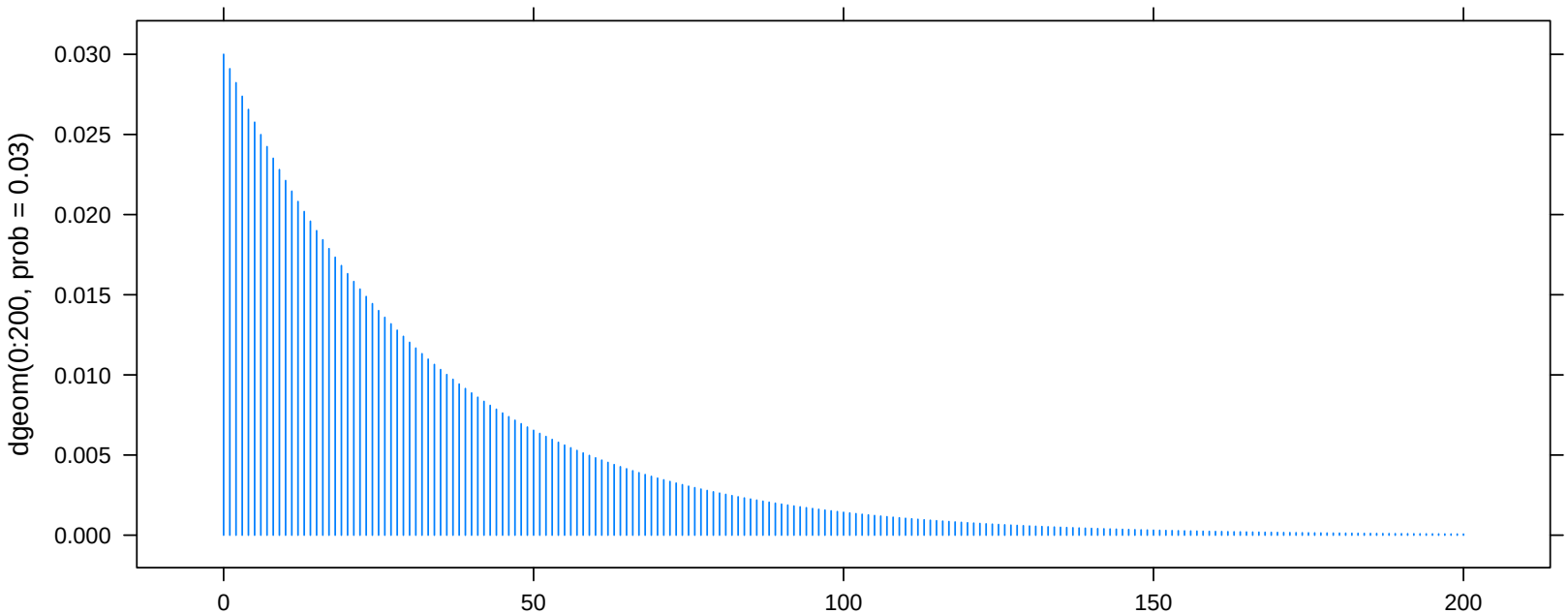
The geometric distribution

- For a Bernoulli process (repeated, independent trials with 0/1 outcomes and constant probability, p, of "success") the binomial distribution arises from setting the number of trials to n and counting the total number of successes.
- The geometric distribution is the distribution of the number of trials before the first success. Be aware that it can be defined in two ways: either count the number of trials until the first success or count the number of failures before the first success. These counts just differ by 1, of course, but you should be careful to check which definition you are using.
- The text counts the total number of trials. The *R* functions `dgeom`, `pgeom`, `qgeom` and `rgeom` are based on the number of failures before the first success.
- In the text's notation, if $\mathcal{Y} \sim \mathrm{Geom}(p)$ then

$$P(\mathcal{Y} = k) = p_k = p(1-p)^{k-1} \quad k = 1, 2, \ldots$$

Example 4.2.13

- We assume that the probability of a defective is 0.03 and check units until a defective is found. Let $\mathcal{Y}$ be the number checked. (Notice that "success" in this case means finding a defective.)
- The expected number of trials is $1/0.03 \approx 33$. The probability distribution of the number of "failures" (non-defectives, in this case) before the first "success" is



Example 4.2.13

- The question in the text is whether the total number of trails is ≤ 3.
- In *R* we need to check if the number of failures before the first success is ≤ 2. We can use the probability mass function, `dgeom`, or the cumulative distribution function, `pgeom`

```
> sum(dgeom(0:2, prob = 0.03))
[1] 0.087327
> pgeom(2, prob = 0.03)
[1] 0.087327
```

Properties of the geometric distribution

- If $\mathcal{Y} \sim \text{Geom}(p)$ (text's definition) then $\text{E}(\mathcal{Y}) = \frac{1}{p}$ and $\text{Var}(\mathcal{Y}) = \frac{1-p}{p^2}$
- In the definition used in *R*, $\text{E}(\mathcal{Y}) = \frac{1}{p} - 1 = \frac{1-p}{p}$ and the variance is the same as above (why?).
- You can check these formulas at Wikipedia, http://en.wikipedia.org/wiki/Geometric_distribution. Information on the other distributions is also available there.
- You can also check it out by creating a very large random sample and taking the sample mean and variance. Then change parameters and do it again.

```
> mean(gsamp <- rgeom(1000000, prob = 0.03)) # should be near 32
[1] 32.28504
> var(gsamp)  # should be near 0.97/0.03^2 = 1077.778
[1] 1074.451
```

Parameter estimation for the geometric

- As before you must be careful if you are counting the total number of trials (including the last trial, which must be a success, by definition) or just the failures before the first success.
- If y is the observed number of trials then $\hat{p} = 1/y$.
- If we have several repetitions of the experiment (recommended) then $\hat{p} = 1/\bar{y}$

Important continuous distributions

- We will use three basic continuous distributions: uniform, exponential and normal (or Gaussian) plus two derived distributions, the lognormal and the Weibull, which is a generalization of the exponential.
- The corresponding *R* functions end in `unif`, `exp`, `norm`, `lnorm` and `weibull`.
- The Wikipedia URLs end in `Uniform_distribution_(continuous)`, `Exponential_distribution`, `Normal_distribution`, `Log_normal_distribution` and `Weibull_distribution`.
- Although we define the distribution by its density, in practice we usually work with the cumulative distribution function.

The uniform distribution

- The density is constant over a finite interval, usually written $[a, b]$. Thus

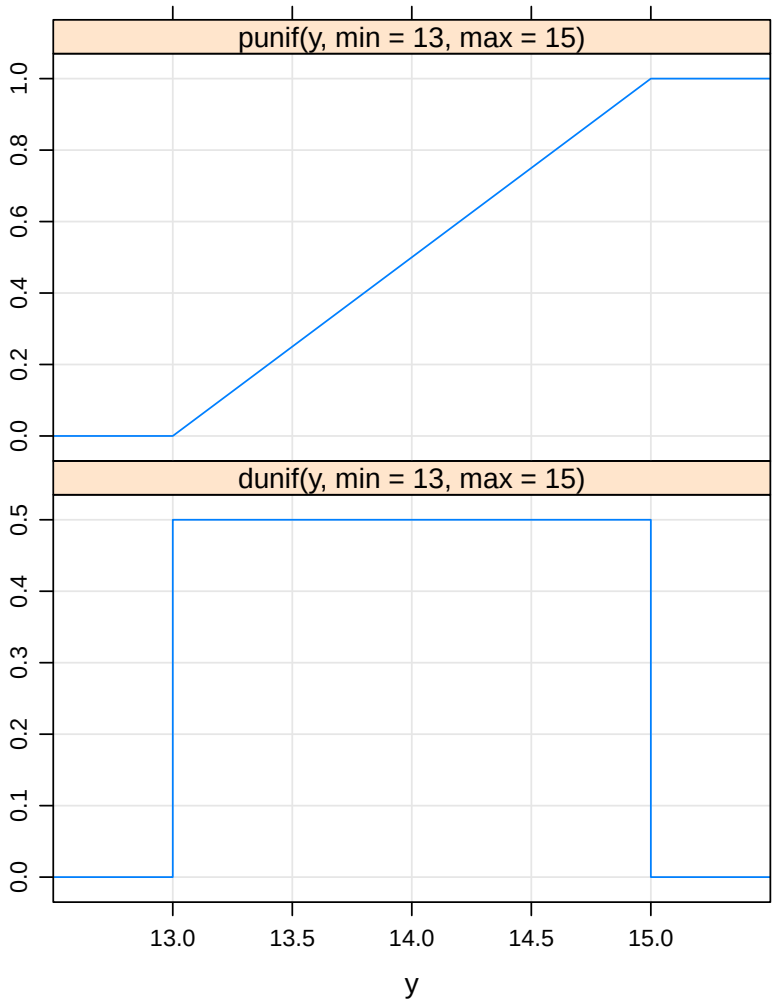
$$f(y) = \begin{cases} \frac{1}{b-a} & a < y < b \\ 0 & \text{otherwise} \end{cases}$$

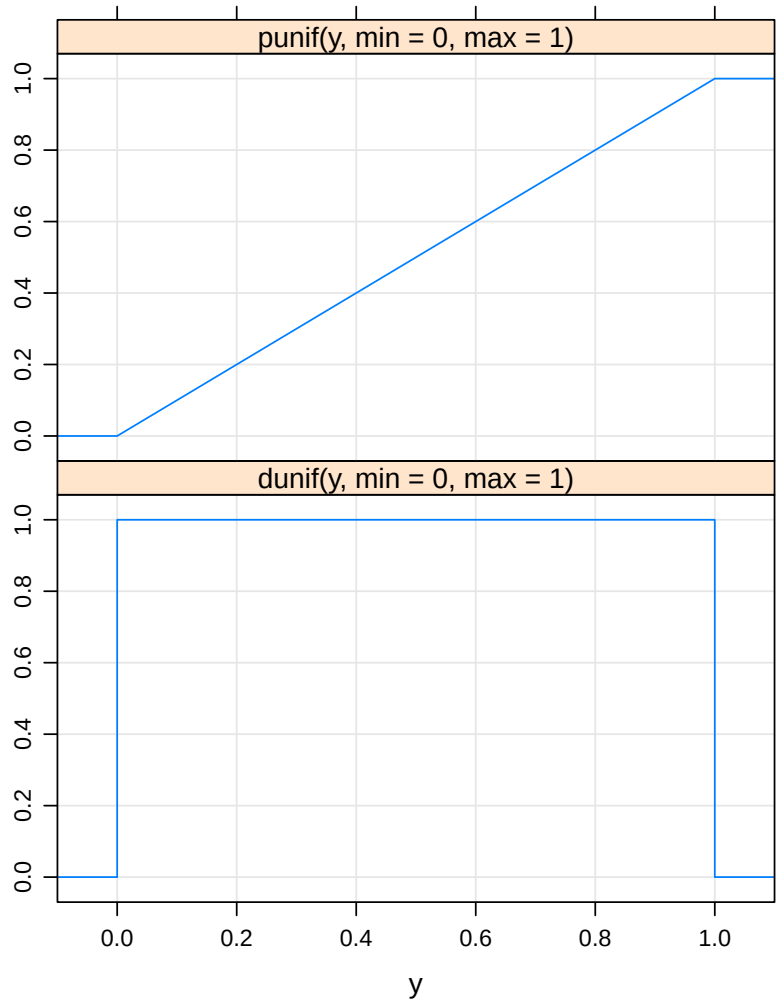
- This density is symmetric about the midpoint of the interval so $\mathrm{E}(\mathcal{Y}) = \frac{a+b}{2}$.
- The cumulative distribution function (cdf) is

$$F(y) = \begin{cases} 0 & y < a \\ \frac{y-a}{b-a} & a \leq y \leq b \\ 1 & y > b \end{cases}$$

- We call $\mathcal{U} \sim \mathrm{Unif}(0, 1)$ the *standard uniform distribution*. It is easy to verify that $\mathrm{Var}(\mathcal{U}) = \int_0^1 \left(u - \frac{1}{2}\right)^2 du = \frac{1}{12}$. For the general uniform distribution, $\mathrm{Var}(\mathcal{Y}) = \frac{(b-a)^2}{12}$.

Example cdf's and pdf's from uniform distributions





The exponential distribution

- A common model for lifetime data (also called *time to failure* or *survival analysis*) is the exponential distribution with density

$$f(y) = \begin{cases} 0 & y \leq 0 \\ \frac{e^{-y/\mu}}{\mu} & y > 0 \end{cases}$$

  where $\mu > 0$ is the expected time to failure.
- This distribution is often written in terms of the failure rate, $\lambda = 1/\mu$, for which $f(y) = \lambda e^{-\lambda y}$, $y > 0$.
- The parameter λ is the same as for the Poisson distribution. Just as the geometric distribution is the number of trials before the first event for a Bernoulli process, the exponential distribution is the length of time before the first event in a Poisson process.

Cumulative distribution function (cdf) for exponential

- The exponential distribution is one of the few cases where the cdf is easier to evaluate than the density.

$$F(y) = \int_{-\infty}^{y} f(u)\,du = \int_{0}^{y} f(u)\,du = 1 - e^{-y/\mu} \quad y > 0$$

- This makes it particularly easy to evaluate probabilities. Note that the *R* functions `pexp`, `dexp`, `qexp` and `rexp` use the parameter $\lambda = \frac{1}{\mu}$, not μ.
- In Example 4.3.4 the lifetimes of compact fluorescent light bulbs are assumed to be exponentially with mean $\mu = 7$ years. The probability a given bulb will last less than 3.5 yr. is

$$P(\mathcal{Y} < 3.5) = 1 - e^{-3.5/7} = 0.39$$

(Recall that for continuous r.v.'s you can interchange $<$ and $\leq$.) In *R* you would use

```
> pexp(3.5, rate = 1/7)
```

```
[1] 0.3934693
```

Example 4.3.5

- The probability that such a light bulb will last more than 6 years can be evaluated in two ways

```
> 1 - pexp(6, rate = 1/7)

[1] 0.4243728

> pexp(6, 1/7, lower = FALSE)

[1] 0.4243728
```

- `lower = FALSE` is used with any cumulative probability function to obtain the probability of exceeding the first argument.
- The probability that such a light bulb lasts between 3.5 and 9 years is evaluated as the difference in the c.d.f. values. I prefer to use the diff function to do this.

```
> pexp(9, 1/7) - pexp(3.5, 1/7)

[1] 0.3300776

> diff(pexp(c(3.5, 9), 1/7))

[1] 0.3300776
```

Properties and parameter estimation

- As indicated by the notation, the expected value of the exponential distribution is the mean, μ. Sometimes this is called the *mean time between failures* or *MTBF*.
- If $\mathcal{Y}$ is written in terms of the rate or intensity, λ, then $\mathrm{E}(\mathcal{Y}) = \frac{1}{\lambda}$.
- The variance is μ^2 or $\frac{1}{\lambda^2}$, hence the standard deviation is $\sigma = \mu$ for an exponential distribution.
- The estimate of μ is $\widehat{\mu} = \bar{y}$. Correspondingly, the estimate of the intensity, λ, is $\widehat{\lambda} = \frac{1}{\bar{y}}$.

The normal or Gaussian distribution

- By far the most important distribution in statistics is the normal or Gaussian distribution, the so-called "bell curve".
- It has two parameters, μ and σ. These turn out to be the mean and the standard deviation, respectively.
- The density for $\mathcal{Y} \sim \mathcal{N}(\mu, \sigma^2)$

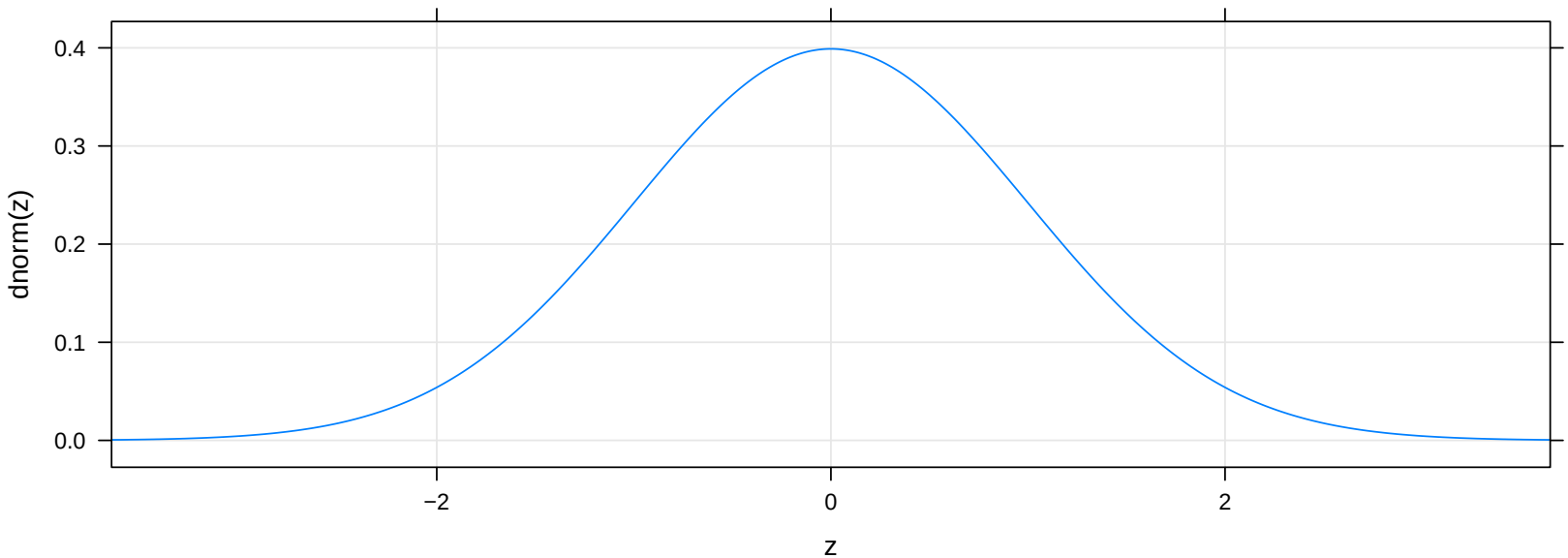
$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

- There is no closed-form expression for the c.d.f., $F(x)$, but good numerical methods for its evaluation are available.
- The text describes methods based on converting from a general normal distribution to the standard normal and referring to tables of that distribution. We can by-pass all that.

The density of the standard normal distribution

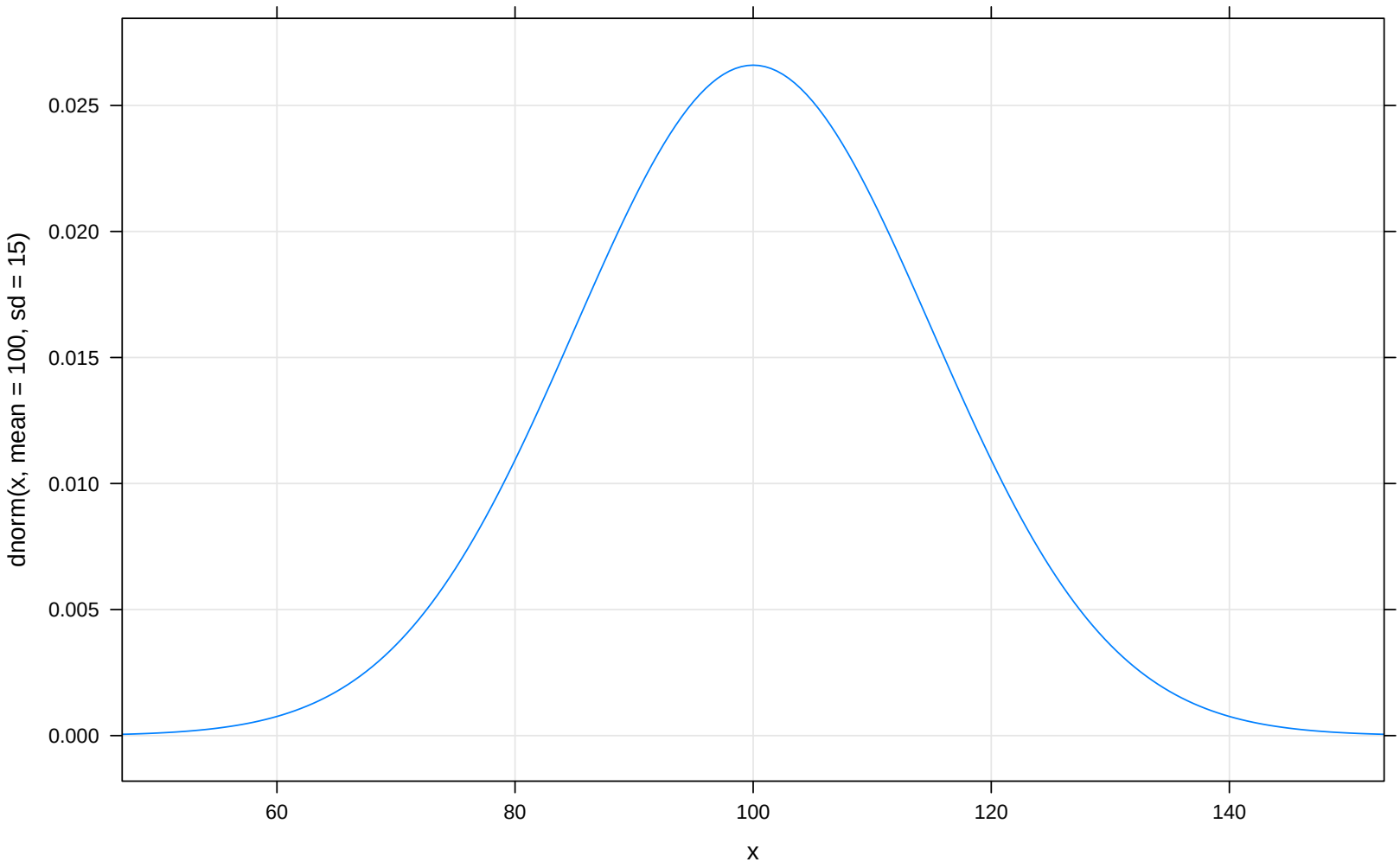
- The *standard* normal distribution is $\mathcal{N}(0, 1)$. That is $\mu = 0$ and $\sigma = 1 = \sigma^2$.
- The parameters of the distribution are specified to the *R* functions `dnorm`, `pnorm`, `qnorm` and `rnorm` as `mean` and `sd`. They default to 0 and 1, respectively.

```
> xyplot(dnorm(z) ~ z, type = c("g","l"))
```



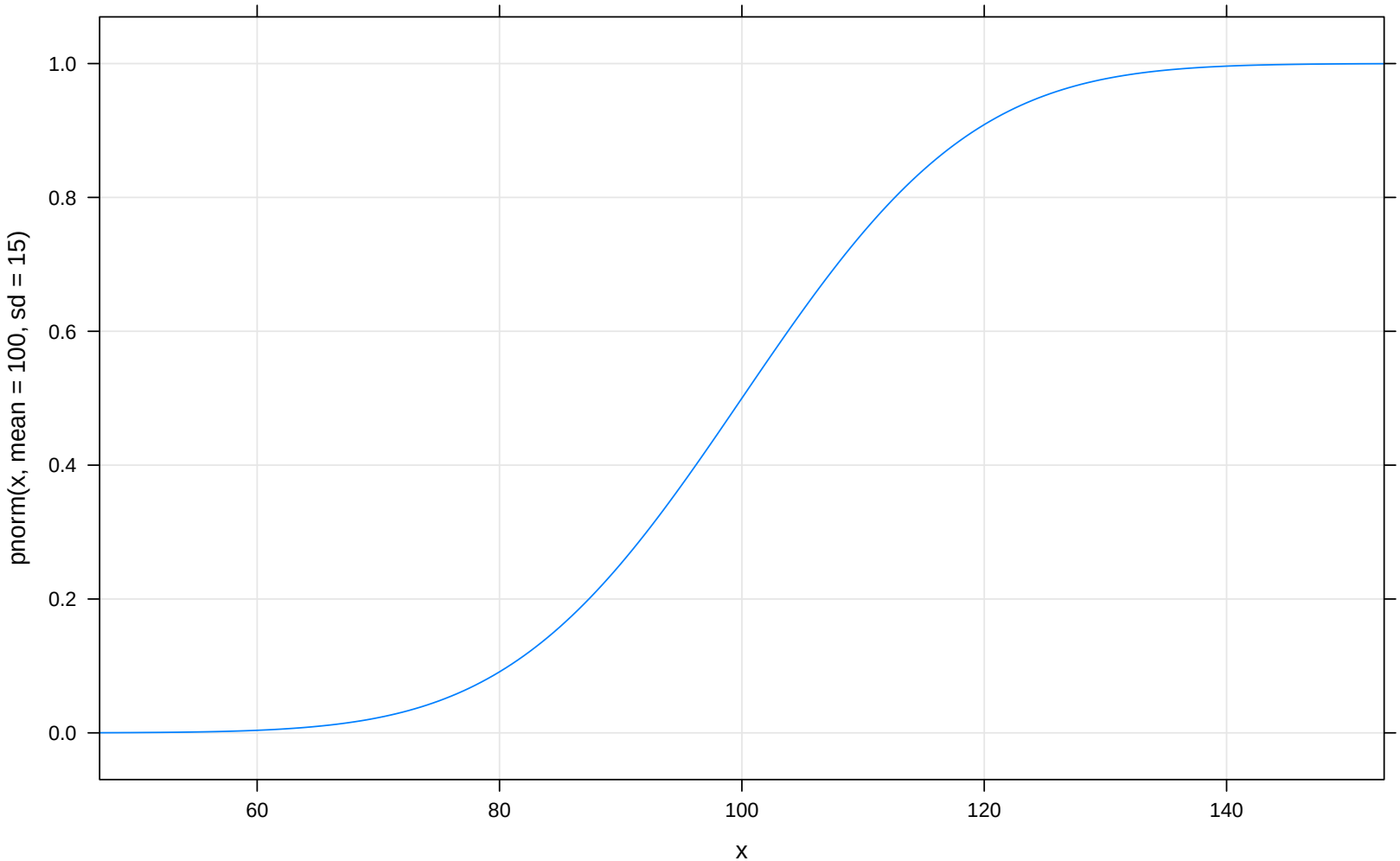
Density of $\mathcal{N}(100, 15^2)$

The IQ scale is normalized to have a mean of 100 and a standard deviation of 15.



Cdf of the Gaussian distribution

The cdf for a Gaussian distribution is sigmoidal (i.e. S-shaped).



Example 4.3.11

- In the example the yields are distributed as $\mathcal{N}(450, 900)$ meaning that $\mu = 450$ and $\sigma = \sqrt{900} = 30$. To evaluate $P(\mathcal{Y} < 500)$, $P(\mathcal{Y} > 550)$ and $P(400 < \mathcal{Y} < 500)$ we use

```
> pnorm(500, 450, 30)
[1] 0.9522096
> 1 - pnorm(550, 450, 30)
[1] 0.0004290603
> diff(pnorm(c(400,500), 450, 30))
[1] 0.9044193
```

Quantiles

- Occasionally we want to find the value of $\mathcal{Y}$ with a specified probabilty to the left of it. This is called a *quantile* of the distribution.
- For example, the organization called Mensa has only one qualification for membership - a candidate's IQ must be in the top 2% of the population. If IQ scores are distributed as $\mathcal{N}\left(100, (15)^2\right)$ then the cutoff score is

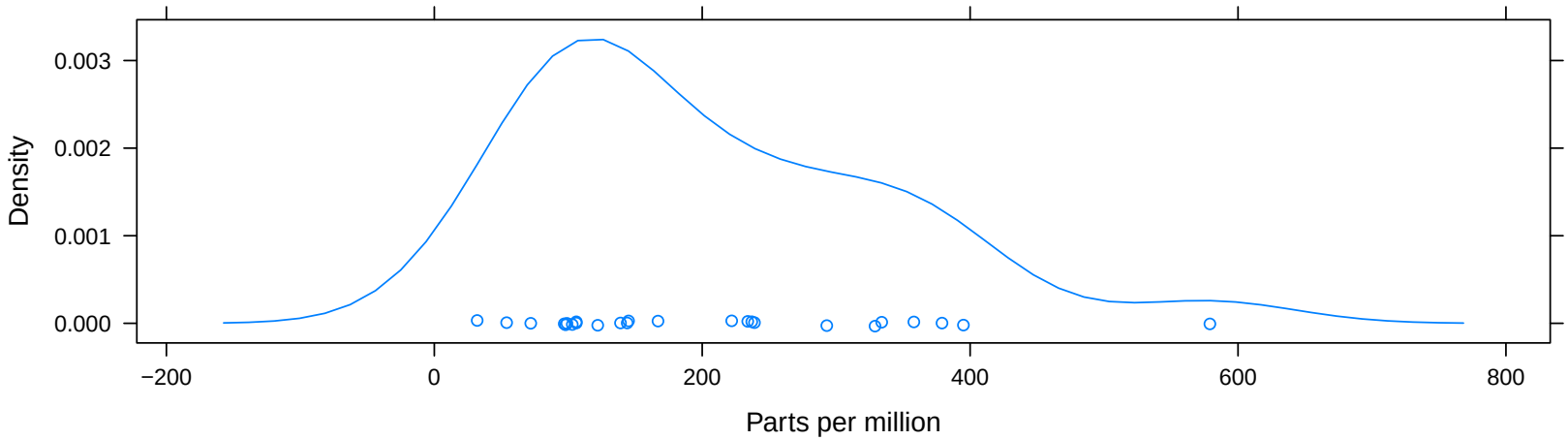
```
> qnorm(0.98, 100, 15)
[1] 130.8062
```

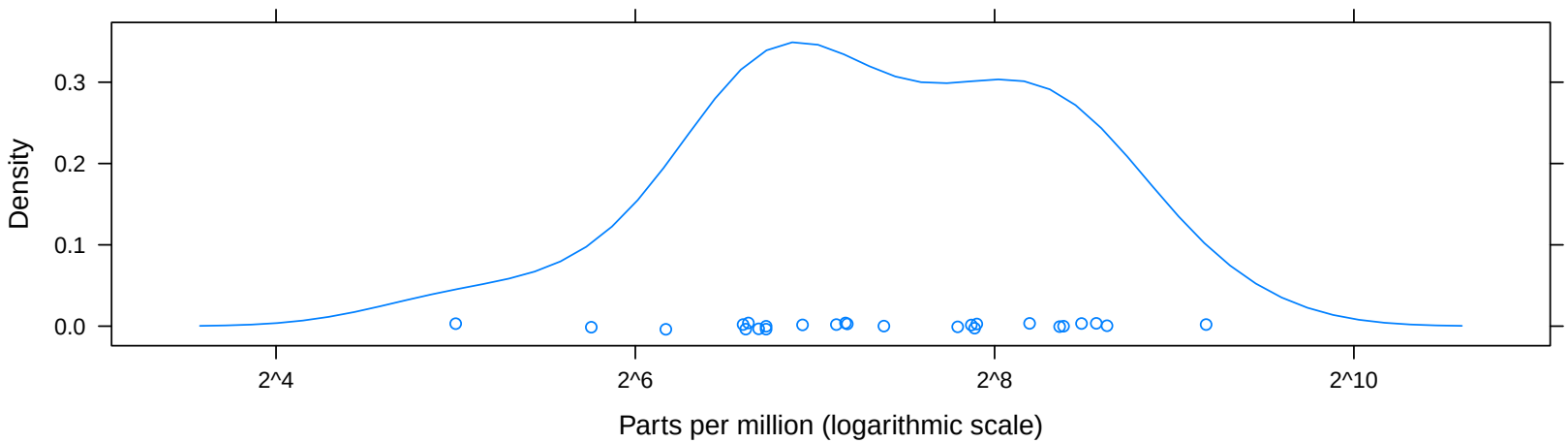
The lognormal distribution

- In many cases the measurements we make must be positive to be physically sensible.
- If the standard deviation is small relative to the mean (like the IQ scores) we can use the normal distribution as a model for the random variable, even though in theory it would give a (very small) positive probability to negative values.
- However, if the standard deviation is not small, compared to the mean, the distribution will usually be skewed. Often the logarithms of the measurements are more like a normal distribution.
- Usually it is simplest to take the logarithms and then use the normal distribution. In some cases there is interest in a model of the results on the original scale, for which we use the lognormal distribution.
- The parameters of the distribution, written μ and σ, are the mean and standard deviation of the *logarithm* of $\mathcal{Y}$. In the *R* functions these are called `meanlog` and `sdlog`.

Example 4.3.20

The `alum` data are measurements of the parts per million of aluminum impurities in the plastics stream at a recycling plant.





Parameter estimates

- The parameter estimates, $\widehat{\mu}$ and $\widehat{\sigma}$ are the sample mean and sample standard deviation of the logarithms of the responses.

```
> with(alum, c(muhat = mean(log(ppm)), sigmahat = sd(log(ppm))))
```

```
    muhat  sigmahat
5.0999591 0.6958257
```

- Given these parameters we can evaluate probabilities either by first taking logarithms and using `pnorm` or by using `plnorm`. If $\mathcal{Y}$ is the impurity level in parts per million then $P(\mathcal{Y} < 100)$ is

```
> plnorm(100, meanlog = 5.0999591, sdlog = 0.6958257)
```

```
[1] 0.2385168
```

```
> pnorm(log(100), mean = 5.0999591, sd = 0.6958257)
```

```
[1] 0.2385168
```

The Weibull distribution

- The Weibull distribution is a generalization of the exponential distribution and is typically used to model the time to failure for systems with an increasing hazard function (i.e. subject to wear-out failures) or with a decreasing hazard function (subject to "break-in" failures).
- If $\mathcal{W}$ has a Weibull distribution with scale parameter β and shape parameter δ then

$$\mathcal{Y} = \left(\frac{\mathcal{W}}{\beta}\right)^{\delta}$$

has a standard exponential distribution (i.e. $\mu = 1 = \lambda$).

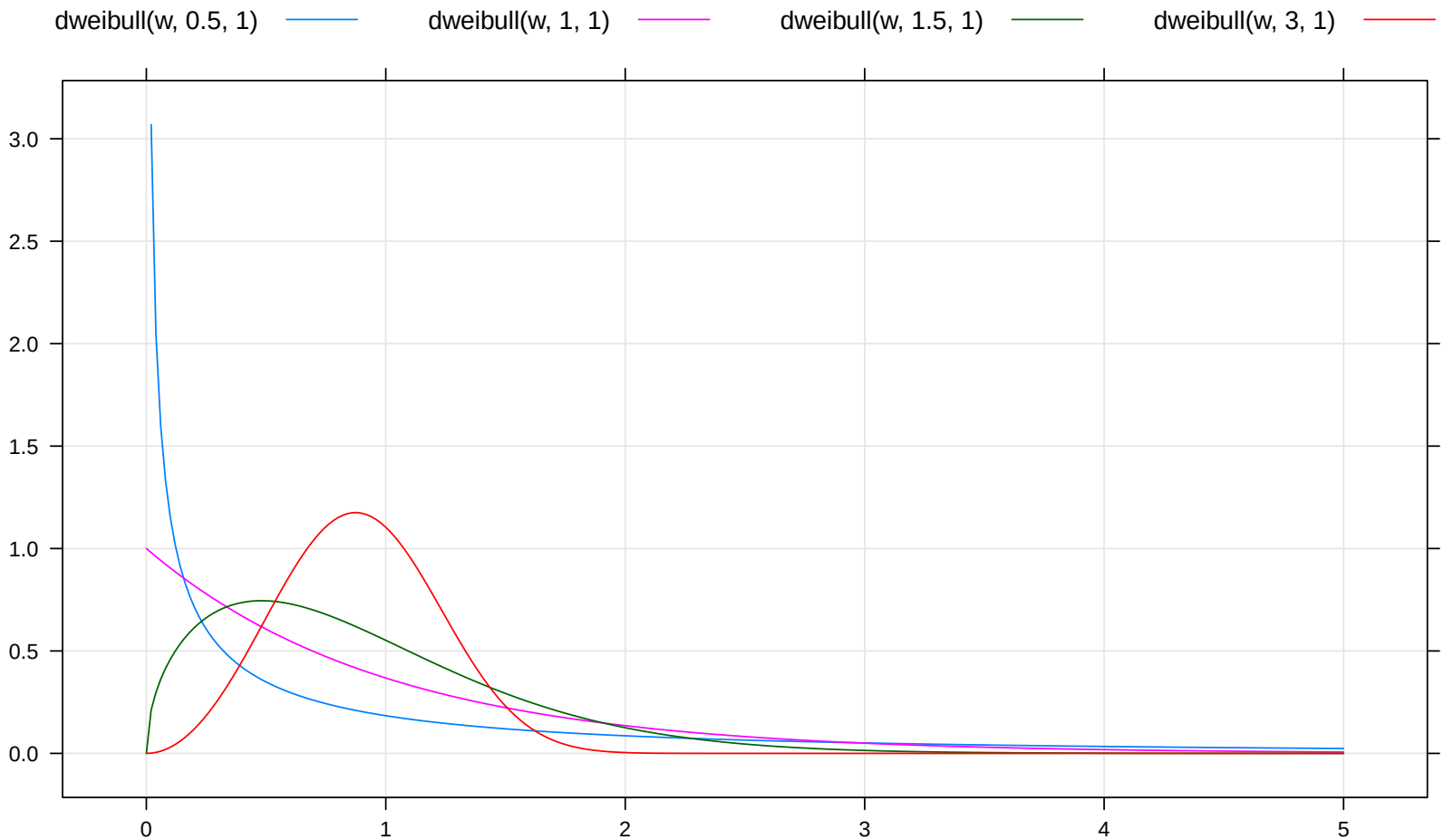
- From these we obtain the cdf

$$F(w) = 1 - e^{-(w/\beta)^{\delta}}$$

which we differentiate to get the density.

Shape parameter and R functions

- When the shape parameter, δ, is 1 the Weibull is the same as the exponential. With $\delta < 1$ it has a decreasing hazard function. With $\delta > 1$ it has an increasing hazard.



Assessing the fit of a distribution

- When we propose a distribution as a model of the random error in measurements it is, like any statistical model, a proposed model and, as George Box famously said, "All models are wrong; some models are useful."
- We should assess if the distribution we are considering is a suitable model and, if not, consider alternatives.
- One of the most powerful methods of assessing a distribution or distribution family as a potential model is through *probability plots* in which the observed quantiles (i.e. the observed sample values in increasing order) are plotted versus the quantiles of the proposed distribution.
- The points form a non-decreasing pattern. We are interested in whether the pattern is close to a straight line.
- When the distribution family incorporates only location and scale parameters we can use the quantiles from the standard form of the distribution, because we are interested only in the pattern and that is not changed by shifting and scaling an

Probability plots (cont'd)

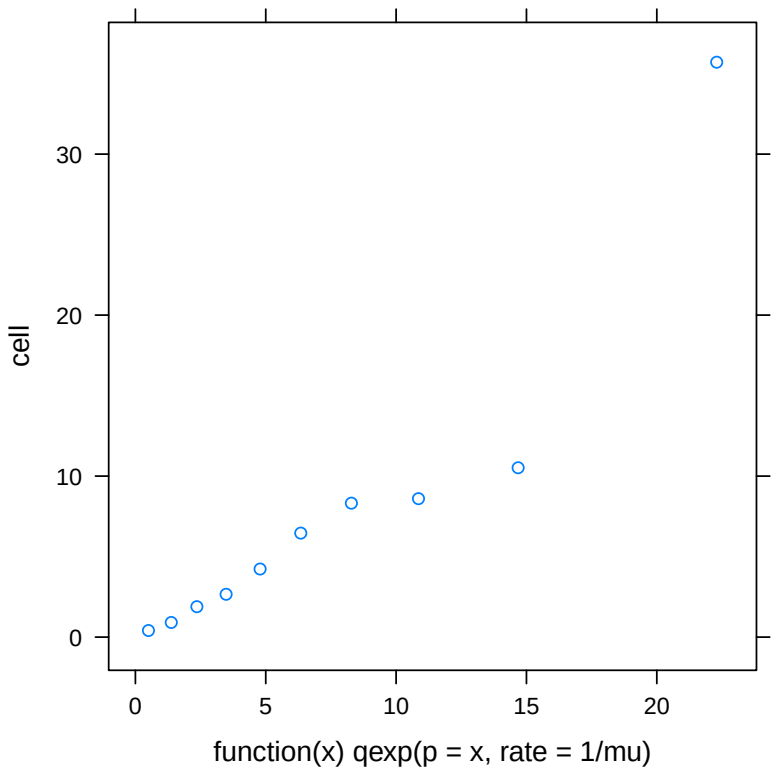
- In general we need to pick a particular distribution (i.e. we must estimate the parameters in the distribution family) to evaluate the theoretical quantiles.
- For the exponential and normal distributions, in which the parameters are shift and scale parameters, we can use the standard form of the distribution.
- Because the exponential distribution has a closed-form and invertible c.d.f. we can easily determine the quantiles.
- We plot the ordered data values, $x_{i)}, i = 1, \ldots, n$ versus $F^{-1}(\frac{i-0.05}{n})$. Notice that we use $\frac{i-0.5}{n}$, not $\frac{i}{n}$. These are the midpoints of n equally spaced intervals from 0 to 1, sometimes called the *probability points*

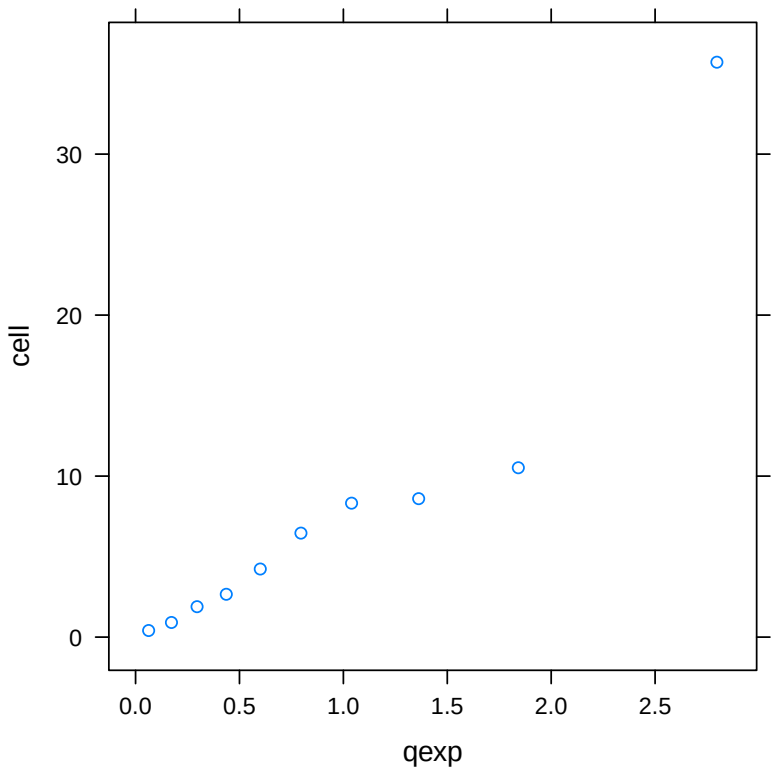
```
> ppoints(20)

 [1] 0.025 0.075 0.125 0.175 0.225 0.275 0.325 0.375 0.425
[10] 0.475 0.525 0.575 0.625 0.675 0.725 0.775 0.825 0.875
[19] 0.925 0.975
```

Example 4.4.1

```
> cell <- c(4.23,1.89,10.52,6.46,8.32,8.6,0.41,0.91,2.66,35.71)
> (mu <- mean(cell))
```

```
[1] 7.971
```





Example 4.41 (cont'd)

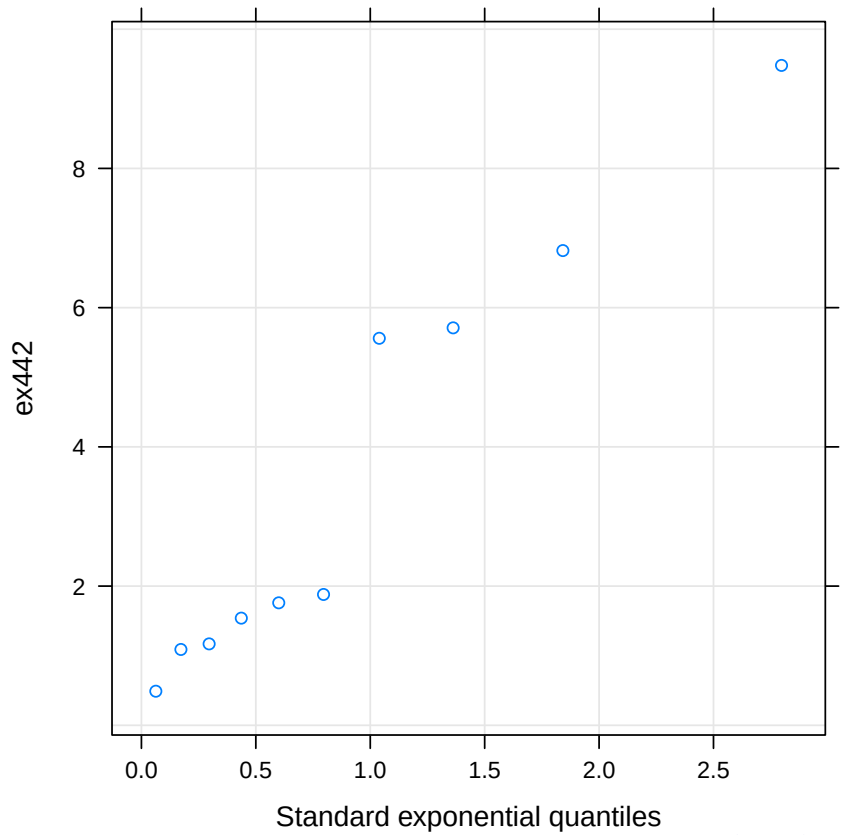
- This example provides (probably fictitious) data on the lifetimes of cellular telephones.
- To obtain a specific exponential distribution we would estimate the parameter μ by the sample mean lifetime (7.971 years) or the parameter λ, the rate, as $1/\widehat{\mu}$.
- The function `qqmath` for drawing a quantile-quantile plot can take a formula/data pair as the first two arguments or the name of a numeric variable. An optional argument called `distribution` (the name can be abbreviated) describes the quantile function to use. Typically this is just the name of a quantile function with the parameters at their defaults.
- These plots were produced by

```
> qqmath(cell, dist = function(x) qexp(x, rate = 1/mu))
> qqmath(cell, dist = qexp)
```

Example 4.4.2

Usually I set the aspect ratio of a qq plot to be 1 so it is easier to recognize a straight line.

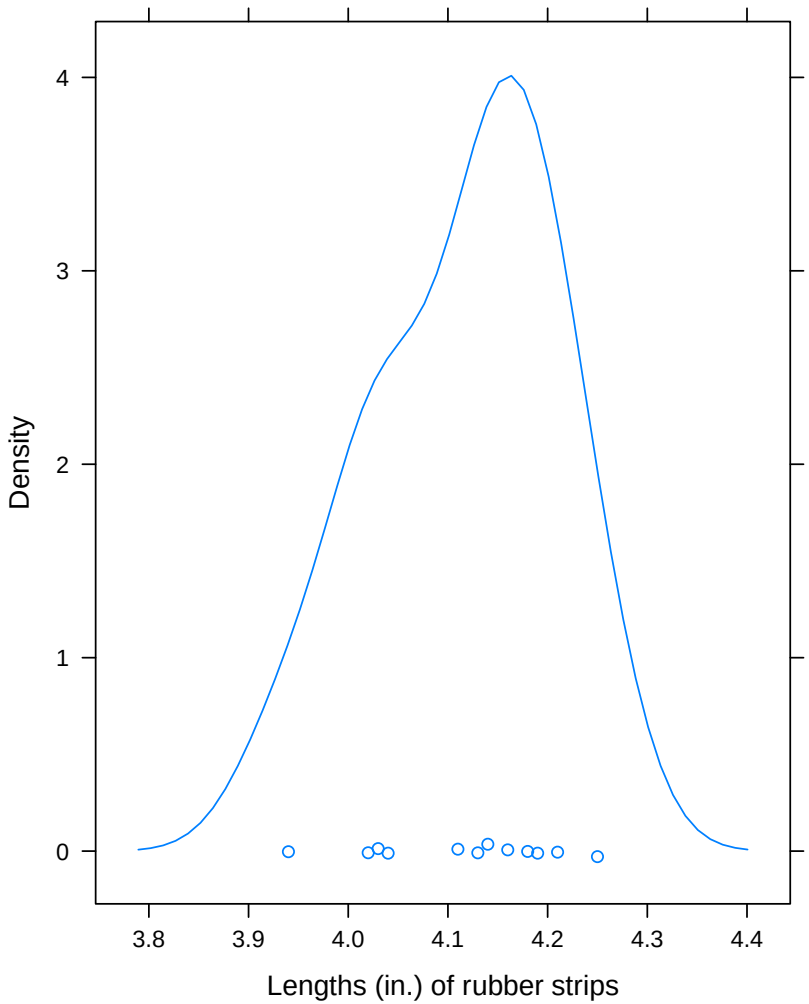
```
> ex442 <- c(1.76,5.71,1.17,0.49,1.09,5.56,6.82,9.48,1.54,1.88)
> qqmath(ex442, dist = qexp, type = c("g","p"), aspect = 1,
+        xlab = "Standard exponential quantiles")
```

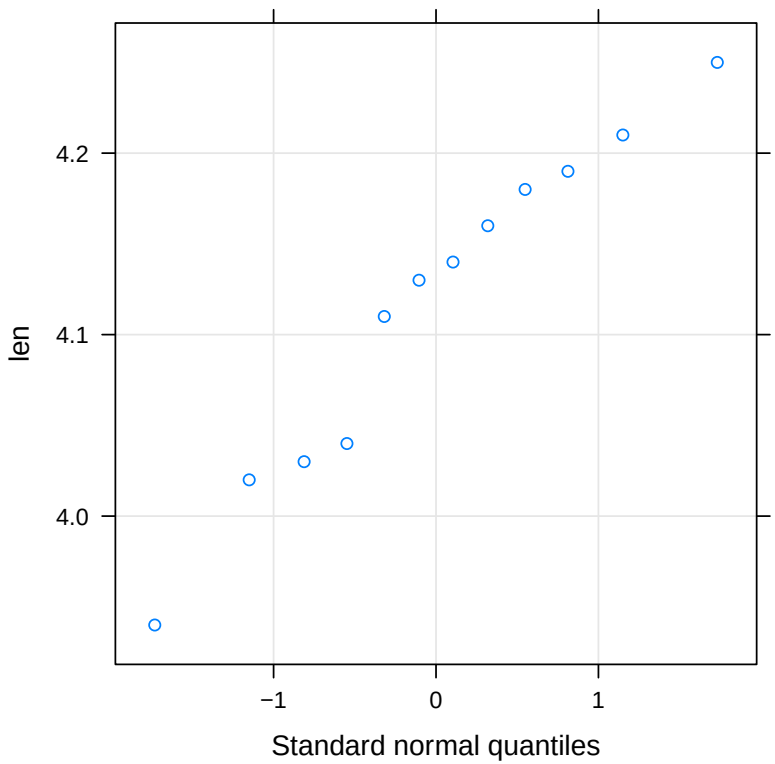


- By far the most important probability plot is the normal probability plot.
- We use it to assess whether a set of observations appears to come from a normal distribution (a reasonably straight increasing line) or is too skewed (a "J-shaped" pattern) or has outliers (points too far down on the left or too far up on the right) or has "heavy tails" (same sort of pattern as outliers).
- We can use the quantiles from the standard normal distribution as the reference. This is the default distribution for the `qqmath` function so we do not need to specify it.
- The combination of an empirical density plot and a normal probability plot is usually the best way to evaluate whether data could have come from a normal distribution.
- The y-intercept of the line is the sample median. If the line is reasonably straight the intercept is close to the sample mean and the slope is close to the standard deviation.

Example 4.4.3

```
> len <- c(4.03,4.04,4.16,4.02,4.18,4.14,4.11,4.13,4.19,3.94,4.2
> qqmath(len, aspect = 1, type = c("g","p"), xlab = "Standard ..
```





Further comments on probability plots

- The text shows examples of fitting straight lines by least squares to the points on a probability plot. Generally this is not considered a good idea because the points at the edges, which are the ones you often want to ignore, are the most influential points in a least squares fit.
- The normal probability plot is the basis for a statistical test of normality called the Shapiro-Wilk test implemented in the `shapiro.test` function. A low p-value (i.e. close to zero) indicates "significant" departures from normality.

```
> shapiro.test(len)

Shapiro-Wilk normality test
data:  len
W = 0.9577, p-value = 0.7508
```

Weibull probability plots

- The Weibull distribution family is a scale/shape family, not a location/scale family.
- We could estimate parameters for the density to create a quantile-quantile plot. An example of estimating the parameters is given in section 4.3. A better general method of parameter estimation, called "maximum likelihood" is implemented in the `fitdistr` function in the `MASS` package.

```
> brks <- tensile$bstrength[1:14] # stored data don't agree with
> library(MASS)
> fitdistr(brks, "weibull")

      shape           scale
    2.0576589    216.4293256
 (  0.4590137) ( 29.5239992)
```

Weibull probability plots (cont'd)

To convert the Weibull distribution from a scale/shape to a location/scale family we take a transformation.

```
> xyplot(log(sort(brks)) ~ log(-log(1-ppoints(14))), aspect = 1)
```