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

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

$$
\begin{aligned}
p_k &\geq 0, \quad \text{all } k \\
\sum_{\text{all possible } k} p_k &= 1
\end{aligned}
$$

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 $f(y) = 0.25, 0 \le y \le 4$ 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

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

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

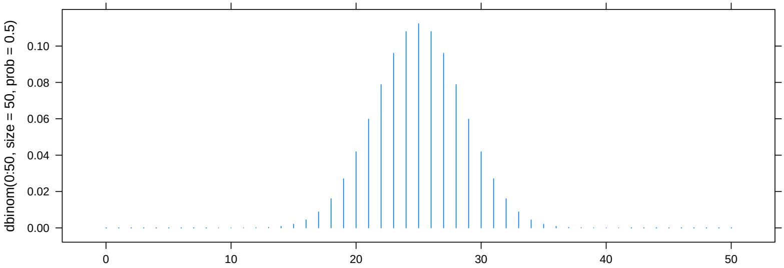
$$\mathrm{E}(\mathcal{Y}) = np$$
$$\mathrm{Var}(\mathcal{Y}) = np(1-p)$$

- 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.

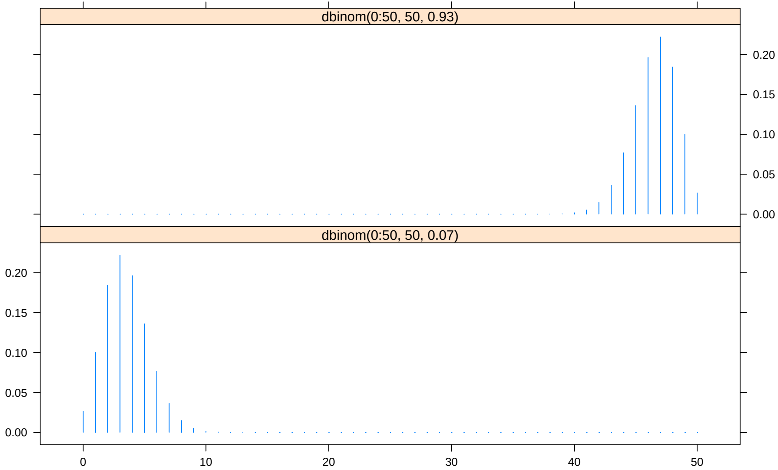
Plotting discrete probability functions

- 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))

[1] 0.4231901
```

- An alternative is to use the cumulative distribution function

```
> ppois(2, lambda = 3)

[1] 0.4231901
```