Chapter 4.3 - Other Named Distributions

> "The statistician cannot evade the responsibility for understanding the processes he applies or recommends." - Sir Ronald A. Fisher

It's not uncommon for scientists to try to do all of their own statistical analyses. Statisticians are expensive and under staffed relative to the number of researchers performing studies everyday. We're also hard to work with considering we're the more sociable versions of applied mathematicians; not an incredibly high bar.

As a result there are many bad statistical analyses riddled throughout academic research and a shocking proportion are considered to be standard practice for their field. Non-normal distributions can fix most of these problematic publications and unfortunately they're not a new concept in statistics.

> "It's a good dissertation topic in Statistics to invent a new probability distribution… completely unimpactful, but you'll get a doctorate." - Anonymous

---

When we ran an experiment of flipping 3 coins 1000 times we went through a series of calculations to derive the mean:

Possible Values	0	1	2	3
Result	147	365	367	121
Weighted Value	0	365	734	363

$$\frac{0 + 365 + 734 + 363}{1000} = \frac{1462}{1000} = 1.462$$

This was done so we could match our result with the one calculated with R:

```r
set.seed(73)
cat("Mean:", mean(rbinom(1000,3,0.5))) # average 1000 iterations of 3 coin tosses
```

```
## Mean: 1.462
```

Which was all a grand effort to prove the expectation from the theoretical probability distribution:

Number of Heads	0	1	2	3
Probability	0.125	0.375	0.375	0.125
Weighted Probability	0.000	0.375	0.750	0.375

$$0.000 + 0.375 + 0.750 + 0.375 = 1.5$$

This is an incredible exercise in justifying where the formula for discrete expectation comes from but it was far from the most efficient method of calculating the expected value. While I may be the author, and thus aware of how this book is written before you've read it, let's check out a neat party trick.

We're going to flip 5 coins 1000 times instead of three and calculate the mean of our results. Before we do this though let's look at the expectation we previously calculated, 1.5.

If the coin is a fair coin (spoiler: it is) then the probability is 1/2, which means 1 out of every 2 tosses *should* land on heads. The expected value we calculated is the result of 3 tosses and claims that 1.5 tosses will come up as heads. When we multiply 3 (the number of coins) by the probability of landing on heads: $3 \times 0.5 = 1.5$. That could be a coincidence so let's verify with our new experiment.

$$5 \times 0.5 = 2.5$$

Number of Heads	0	1	2	3	4	5
Probability	0.03125	0.15625	0.3125	0.3125	0.15625	0.03125
Weighted Probability	0.00000	0.15625	0.62500	0.93750	0.62500	0.15625

$$0.00000 + 0.15625 + 0.62500 + 0.93750 + 0.62500 + 0.15625 = 2.5$$

```r
set.seed(73)
cat("Mean:", mean(rbinom(1000,5,0.5))) # average 1000 iterations of 5 coin tosses
```

```
## Mean: 2.474
```

We'll see this trend no matter how far we go. The expected value for 50 coin tosses is 25:

$$50 \times 0.5 = 25$$

```r
set.seed(73)
cat("Mean:", mean(rbinom(1000,50,0.5))) # average 1000 iterations of 50 coin tosses
```

```
## Mean: 24.859
```

We're also able to apply a formula to the variance. The variance for 50 coins should be $50 \times 0.5(1 - 0.5) = 12.5$, we'll simulate 10000 since variance tends to be less accurate with smaller samples:

```r
set.seed(73)
cat("Variance:", var(rbinom(10000,50,0.5))) # average 10000 iterations of 50 coin tosses
```

```
## Variance: 12.33375
```

Why do these calculations work so well?

Flipping a coin is a classic example of an independent, discrete, *binary* outcome. So classic, in fact, that we can put a name to the probability distribution associated with it.

Discrete Distributions

Whenever we flip a coin more than once we might like to *believe* that the previous flips had something to do with the outcome of the last one, but we know this isn't true. The probabilities attached can be a little confusing though.

The code below simulates 3 fair coin flips and records 1 as heads and 0 as tails. We'll check each flip and discuss the probability of its occurrence.

```
set.seed(73)
coins=ifelse(rbinom(3,1,0.5)==1,"Heads","Tails")
coins[1]
```

```
## [1] "Tails"
```

We know that this was a 0.5 probability outcome, canonically referred to as 50/50.

```
coins[2]
```

```
## [1] "Tails"
```

The probability that we have 2 tails in 2 flips isn't 50/50 though:

Flip 1	Flip 2
H	H
H	T
T	H
T	T

The probability of this specific realization of the random variable is 1/4 or 0.25 and the probability that we see a *third* tails is 1/8 or 0.125, *clearly* we're unlikely to see a third tails right?

```
coins[3]
```

```
## [1] "Heads"
```

This is where the gambler's fallacy begins, but that's not the point of this discussion. The point is that while the *cumulative* probabilities may suggest that seeing a third tails was unlikely the actual outcome was still 50/50. The coin flips are **independent** from one another— landing on tails doesn't change the chance of landing on it again.

Each coin flip had a single result with two possible outcomes, heads or tails. This combination of independent and binary outcomes is referred to as a **Bernoulli trial**.

Bernoulli trial: A random experiment with two possible outcomes, success or failure.

We can redefine what success and failure look like but we're still left with two possible outcomes each time. When describing the distribution of Bernoulli trials we use the **Bernoulli distribution** which is a **discrete** probability distribution with one parameter, p, the probability of success.

$$X \sim \text{Bern}(p)$$

$$P(X = x) = \begin{cases} p & \text{if} \;\; x = 1 \\ 1 - p & \text{if} \;\; x = 0 \end{cases}$$

The **expected value** of a Bernoulli random variables is $\mathbb{E}X = p$ and the **variance** is $\mathbb{V}X = p(1-p)$. These formula are simple because the events we're modeling are simple. The Bernoulli distribution only refers to the *individual trials* not the summation of multiple trials. So for the purpose of checking the probability of a single coin flip, the Bernoulli distribution is the right choice. If we wanted to check the probability of multiple coin flips we would need to use the **binomial distribution**.

Binomial distribution: The number of successes from n many trials with probability of success p.

The Bernoulli distribution is a special case of the binomial distribution where $n = 1$. Because of this we can think of the binomial distribution as the summation of multiple Bernoulli trials.

$$X \sim \text{Binom}(n, p)$$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

The function $\binom{n}{x}$ is read as "*n choose x*", we refer to this as the *binomial coefficient*:

$$\frac{n!}{x!(n-k)!} \quad , \quad \text{where } n! = n \times (n-1) \times (n-2) \times ... \times 3 \times 2 \times 1$$

What the binomial coefficient is calculating is the number of *combinations* (so order doesn't matter) that can be chosen from x.

The **expectation** and **variance** of the binomial distribution are:

$$\mathbb{E}X = np \qquad \mathbb{V}X = np(1-p)$$

We can see how this is just an expansion on the Bernoulli expectation and variance using the parameter n.

The binomial distribution is particularly useful when working with infectious disease since all illness is the result of a binary outcome, sick or healthy. If we were to collect a sample of 10000 bats to record positive or negative status for rabies, a disease with a prevalence of less than 1% among bats, we would expect to see somewhere around < 100 bats test positive.

The binomial distribution is often disregarded because of how easy it is to record enough trials to produce normally distributed data, as we've shown with previous examples. But it's application allows us to avoid waiting until we've acquire massive sample sizes to begin analyzing data.

When sampling for bats from a cave to determine the prevalence of rabies we wouldn't expect any one cave to have the same number of bats as another. We can consider the number of bats in each cave to be its own discrete random variable but we can't have those counts play by the same rules as Bernoulli trials. The probability of observing 7000 bats versus 10000 bats could be modeled as binary events but we would be missing a lot of information between those two outcomes. If we treated each count as an isolated outcome we'd still have the issue of each outcome having *near zero* probability.

This is where the **poisson distribution** enters the picture.

Poisson distribution: The number of events in a fixed interval of time with a constant mean.

The poisson distribution is described fully by its one parameter, λ, which is both the mean and variance of the distribution.

$$X \sim \text{Pois}(\lambda)$$

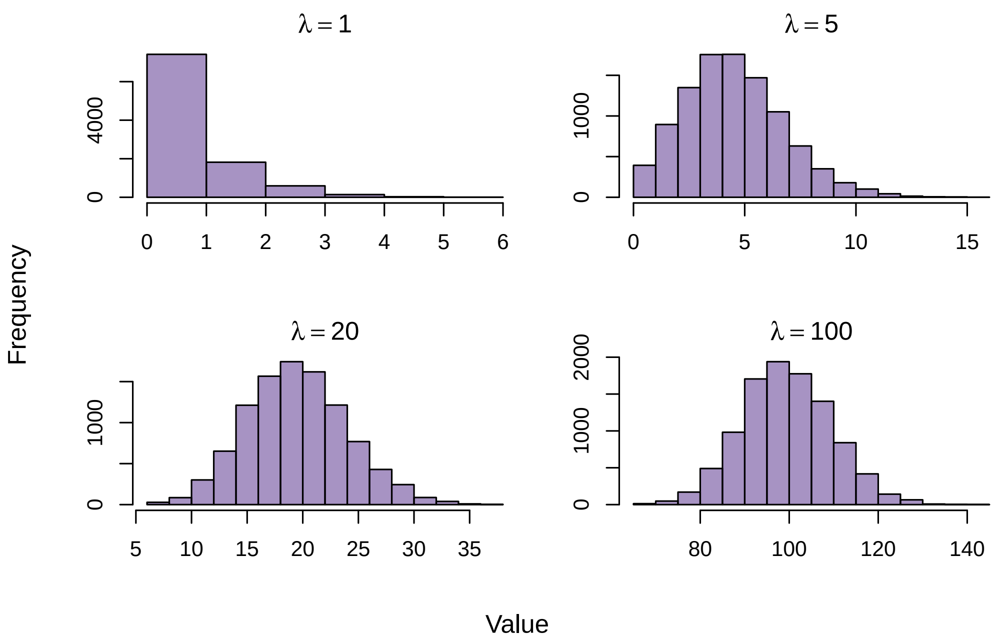
$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

We can use this distribution to model **count data**, things like the number of fish in a pond (which was the first application of this distribution, who's name is also French for 'fish') or bats in a cave.

The most useful feature of this distribution is that its **expectation** and **variance** are the same as its one parameter:

$$\mathbb{E}X = \mathbb{V}X = \lambda$$

This makes the curve of poisson distributed data skewed to the left at low values of λ and look very similar to a normal at higher values:



Continuous Distributions

The Bernoulli distribution is the foundation of all discrete probability distributions, everything can be traced back to this "root distribution". The root of all continuous distributions is the **uniform distribution**.

Uniform distribution: An experiment with equi-probable outcomes enclosed in a defined boundary.

The concept behind this distribution is straight forward: we define a *continuous* interval, say $[1, 5]$, and select a value from that interval at random.

```
set.seed(73)
runif(1,1,5)
```

```
## [1] 2.769348
```

We've discussed this distribution prior because of how simplistic it is. The curve is a flat line and its use cases are countless. A uniform random variable is completely described by its interval parameters. The lower bound, a, and the upper bound, b.

$$X \sim \text{Unif}(a, b)$$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for} \quad a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

The **expectation** of the uniform is $(a + b)/2$ and the **variance** is $(a + b)^2/12$.

This distribution is very prevalent in mathematical statistics due to it being analytically simple. It's also the core of all random number generation in programs like R since it is essentially a perfectly random selection method.

When we modeled the number of rabies positive bats using a binomial distribution we made the assumption that the probability was **fixed**, meaning it's unchanging. If we instead assumed the probability, p, to be a random variable we would need a distribution that's bound between 0 and 1.

A good candidate for this is the standard uniform distribution, a uniform bound between 0 and 1, but it would be helpful to be able to *weight* the distribution to prefer certain ranges of probabilities over others. For this, we use the **beta distribution**.

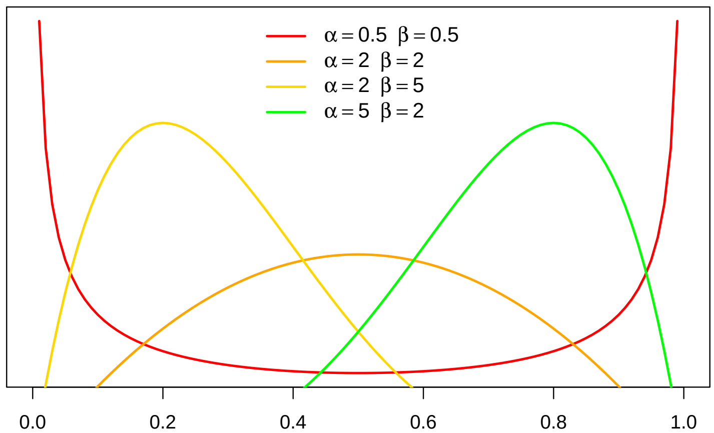
Beta distribution: A continuous distribution defined on the interval $[0, 1]$.

$$X \sim \text{Beta}(\alpha, \beta)$$

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1} \quad , \quad \text{where } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^-t \; dt$$

While I'm showing the PDFs for all of these distributions we don't need to trouble ourselves with them beyond observing their general structure. The application of these functions is a very advanced topic; as the "Gamma function" (Γ) shows.

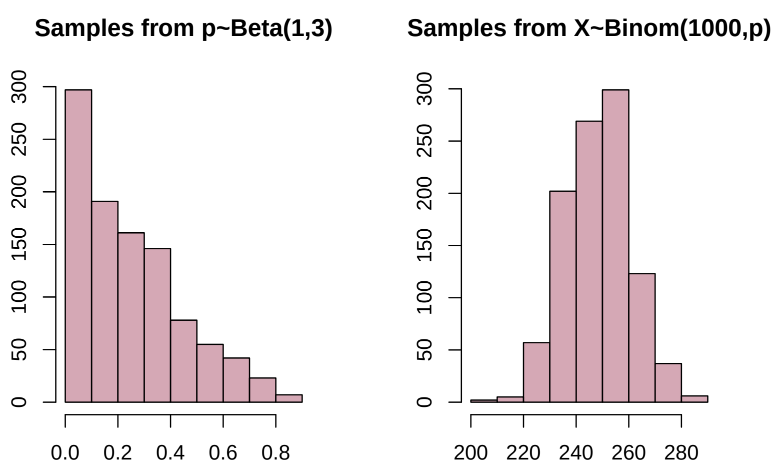
The beta distribution has two parameters, α and β, which control the *shape* of the distribution curve:



The **expectation** and **variance** of the beta distribution are:

$$\mathbb{E}X = \frac{\alpha}{\alpha + \beta} \qquad \mathbb{V}X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The beta distribution is obviously the go-to for any estimation of proportions or probabilities but it's also quite useful when used to model probability parameters of other distributions. This is another one of those topics that could be (and is) an entire book on its own. For now we can wet our whistle with the idea of a binomial process with a variable probability of success:



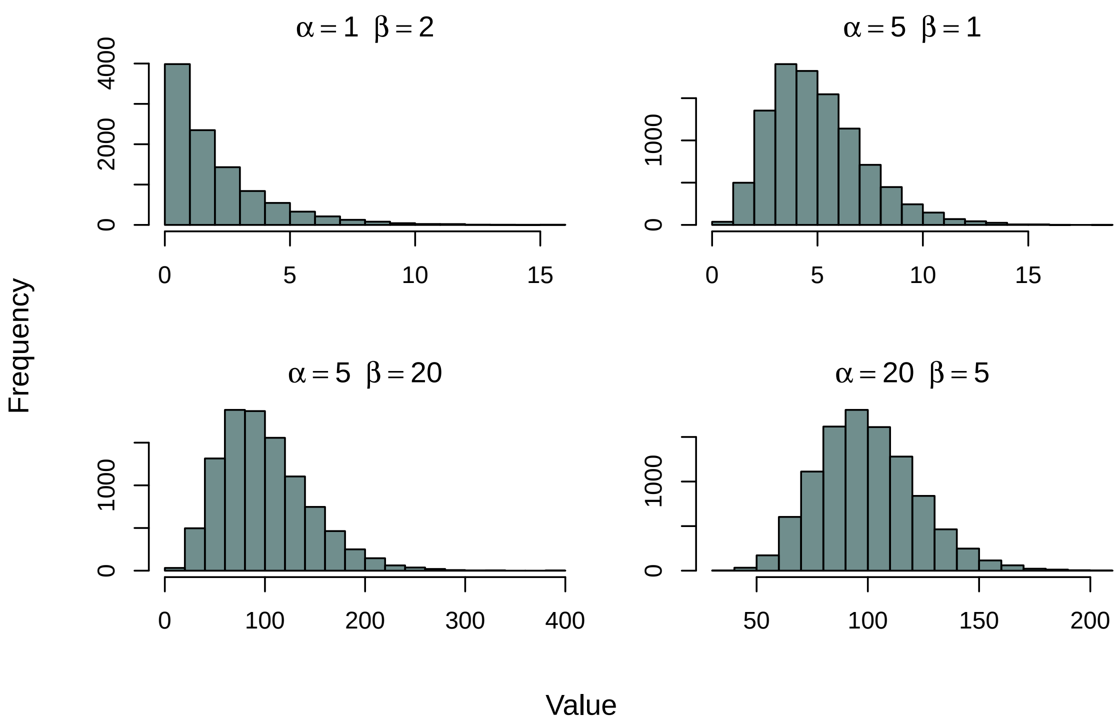
If we're met with data that's continuous but *strictly positive* ($X > 0$) the normal won't always cut it. While the normal is perfectly fine for strictly positive values when μ is very large and σ is very small, what do we do when our values are close to 0 but can never be less than it? One of our options is the **gamma distribution**.

Gamma distribution: A two-parameter, exponential distribution with support on the interval $[0, \infty)$.

$$X \sim \text{Gamma}(\alpha, \beta)$$

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$$

The parameter α controls the shape of the gamma curve and β controls the scale.



This may be a statistics textbook but its focus is on statistics *as it applies to biological scientists*. There are numerous applications of these functions, their parameters, and their moments. But *our* focus needs to be on identifying distributions that fit the data because that's what will allow us to select the proper method for analyzing it.