Statistics for Ecologists

Olivier Gimenez

October 2020

Who's that guy?!

Olivier Gimenez

- Senior scientist at CNRS, in Montpellier - France.
- Trained as a statistician
- Soon attracted by the bright side of ecology
- Working at the interface of animal demography, statistical modelling and social sciences
- More on https://oliviergimenez.github.io/
- Twitter @oaggimenez

Acknowledgments

Acknowledgments

- Sean Anderson, Ben Bolker, Jason Matthiopoulos, David Miller, Denis Réale and Francisco Rodriguez-Sanchez for sharing their courses material

This Class

Slides, R codes, data and practicals

- I used `R`, and `RStudio` is your friend
- I also used `R Markdown` to write reproducible documents (slides/exercises)
- All material is available on GitHub https://github.com/oliviergimenez/statistics-for-ecologists-Master-courses
- Check out the files `gimenez_lectures.R` and `gimenez_practicals.R`
- You will need the following R packages: `arm`, `bbmle`, `broom`, `dplyr`, `effects`, `lme4`, `mgcv`, `MuMIn`, `R2jags`, `tibble`, `visreg`

On our plate

- Distributions and likelihoods
- Hypothesis testing and multimodel inference
- Introduction to Bayesian inference
- Generalized Linear Models (GLMs)
- Generalized Additive Models (GAMs)
- Mixed Effect Models

On our plate

- Distributions and likelihoods
- Hypothesis testing and multimodel inference
- Introduction to Bayesian inference
- Generalized Linear Models (GLMs)
- Generalized Additive Models (GAMs)
- Mixed Effect Models

Distributions and likelihoods

Distributions

- What for?
- Conceptual models, bearing in mind that:

  All models are wrong, but some are useful (G.E.P. Cox, 1976)

- Either represent how the world works
- Or capture the behavior of a statistic under some null hypothesis we'd like to test
- Discrete or continuous

Discrete distributions

Bernoulli distribution

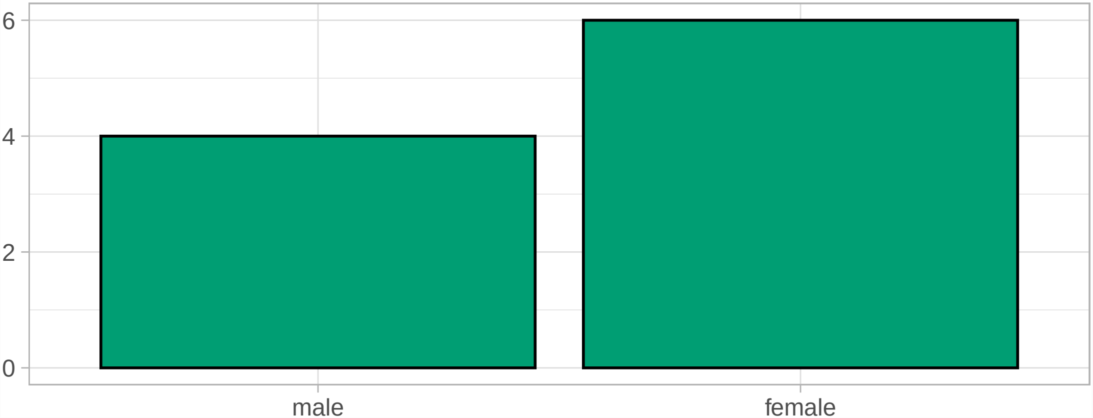
Context: A single trial with two outcomes, success/failure

$X \sim \text{Bern}(p)$ with p probability of having a success

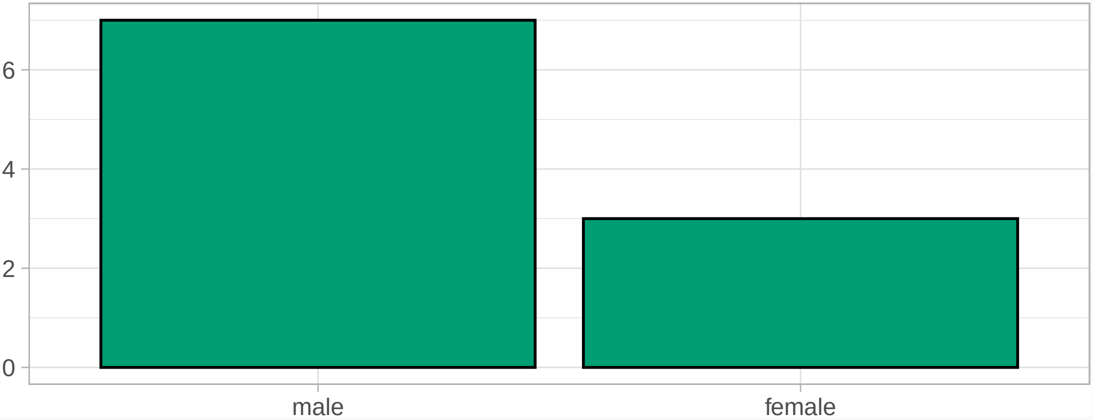
x	$P(X = x)$
1	p
0	$1 - p$

Example: X is the random variable *being born a female*

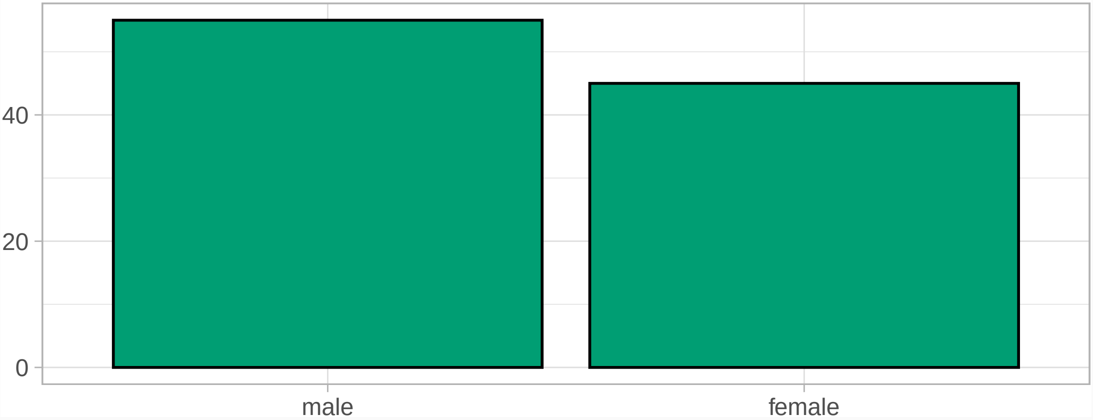
Ten Bernoulli trials with $p = 0.5$



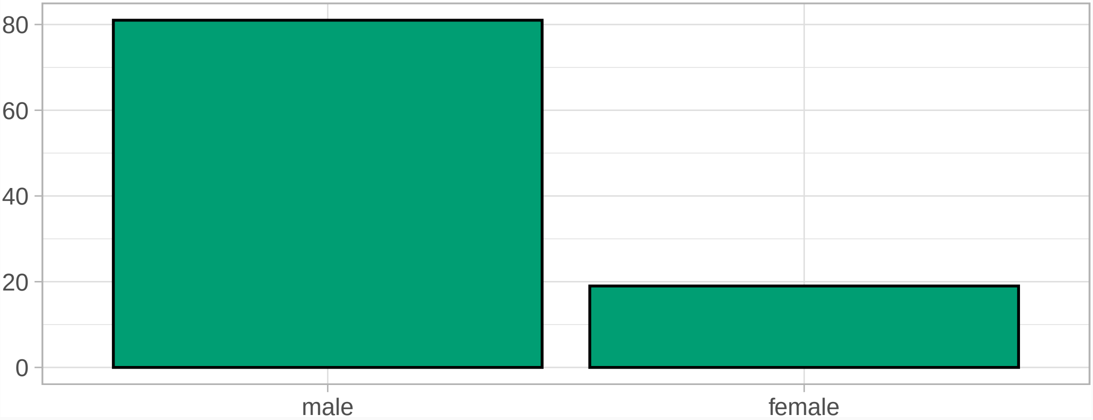
Ten Bernoulli trials with $p = 0.5$, again



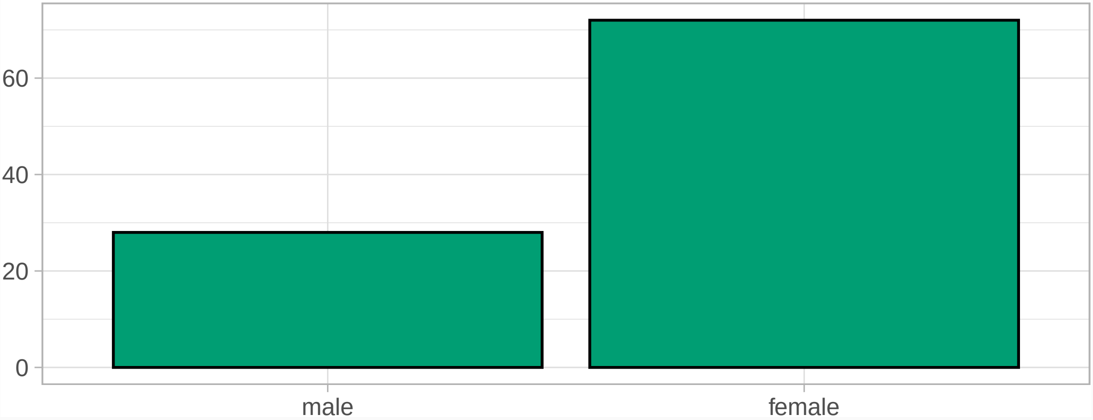
Hundred Bernoulli trials with $p = 0.5$



Hundred Bernoulli trials with $p = 0.2$



Hundred Bernoulli trials with $p = 0.8$



Summary: Bernoulli distribution

- **notation**: $X \sim \text{Bern}(p)$
- **range**: discrete, $x = 0, 1$
- **distribution**: $P(X = x) = p^x (1 - p)^{1-x}$
- **parameters**: p is the probability of success
- **mean**: p
- **variance**: $p(1 - p)$

Binomial distribution

Context: Total number of successes from a fixed number of independent Bernoulli trials, all with same probability of success

$X \sim \text{Bin}(N, p)$ with p probability of having a success and N number of trials

$$P(X = x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} = \binom{N}{x} p^x (1-p)^{N-x}$$

Example: X is the random variable *number of heads in a series of coin flipping*

Binomial distribution

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

x	$P(X = x)$
0	$(1-p)^N$
1	$Np(1-p)^{N-1}$
...	...
N	p^N

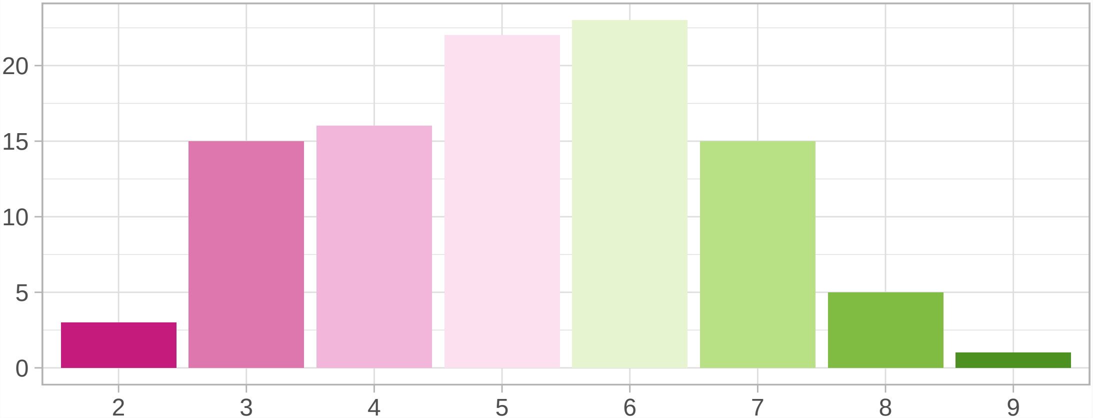
Binomial distribution

x	$P(X = x)$
0	$(1-p)^N$
1	$Np(1-p)^{N-1}$
...	...
N	p^N

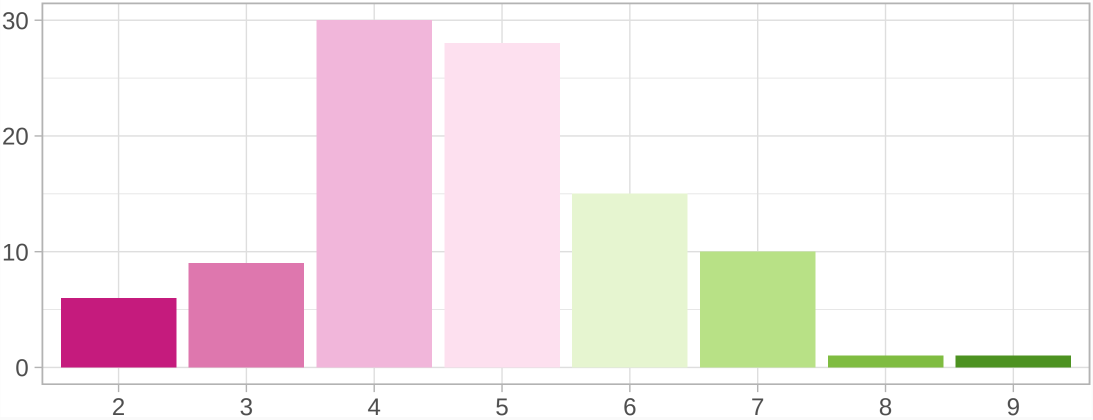
Fortunately, `R` has this pre-programmed

```
dbinom(x = 1, size = 10, prob = 0.5) # equals 10*0.5*(1-0.5)^(10-1)
## [1] 0.009765625
```

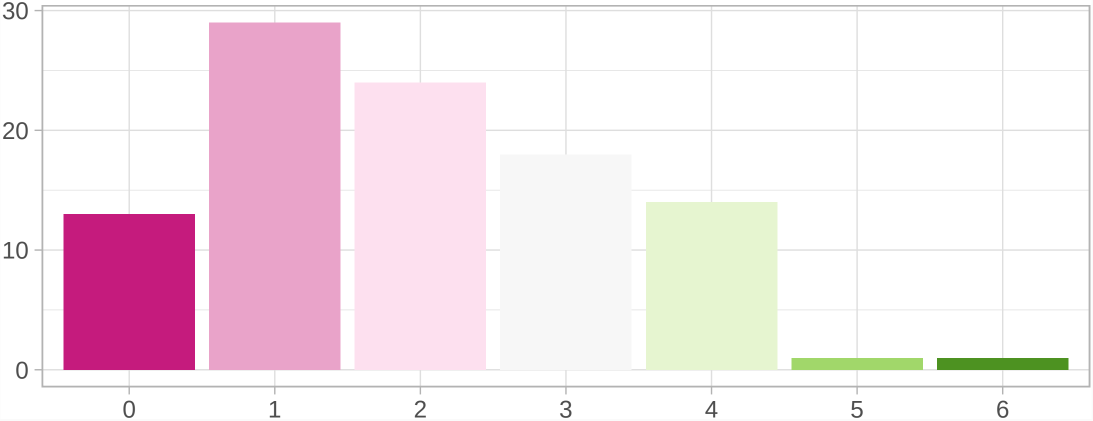
Hundred Binomial trials with $N = 10$ and $p = 0.5$



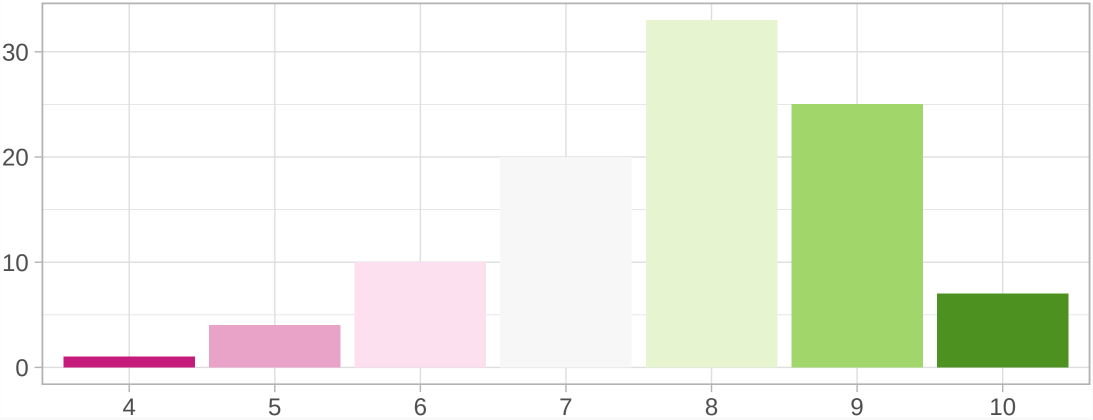
Hundred Binomial trials with $N = 10$ and $p = 0.5$, again



Hundred Binomial trials with $N = 10$ and $p = 0.2$



Hundred Binomial trials with $N = 10$ and $p = 0.8$



Playing around with probabilities

- Let's say $X \sim \text{Bin}(N = 10, p = 0.5)$ is a random variable counting the number of males

Playing around with probabilities

- Let's say $X \sim \text{Bin}(N = 10, p = 0.5)$ is a random variable counting the number of males
- What is the probability of having at most 2 males?

Playing around with probabilities

- Let's say $X \sim \text{Bin}(N = 10, p = 0.5)$ is a random variable counting the number of males
- What is the probability of having at most 2 males?
- $P(X \leq 2) = P(X = 0) + P(X = 1)$

Playing around with probabilities

- Let's say $X \sim \text{Bin}(N = 10, p = 0.5)$ is a random variable counting the number of males
- What is the probability of having at most 2 males?
- $P(X \leq 2) = P(X = 0) + P(X = 1)$
- How to compute this in R?

Playing around with probabilities

- Let's say $X \sim \text{Bin}(N = 10, p = 0.5)$ is a random variable counting the number of males
- What is the probability of having at most 2 males?
- $P(X \leq 2) = P(X = 0) + P(X = 1)$
- How to compute this in R?
- dbinom(x=0,size=10,prob=0.5) + dbinom(x=1,size=10,prob=0.5)

Summary: Binomial distribution

- **notation**: $X \sim \text{Bin}(N, p)$
  - **range**: discrete, $0 \leq x \leq N$
  - **distribution**: $P(X = x) = \binom{N}{x} p^x (1-p)^{1-x}$
  - **parameters**: p the probability of success, and N the number of trials
- **mean**: Np
  - **variance**: $Np(1-p)$
  - **in R**: `rbinom`, `dbinom`

Poisson distribution

Context: Number of occurrences of an event over a given unit of space or time.

$X \sim \text{Poisson}(\lambda)$ with λ expected number of occurrences

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

Example: X is the random variable *number of birds counted on a colony during the breeding season*

Poisson distribution

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

x	$P(X = x)$
0	$e^{-\lambda}$
1	$\lambda e^{-\lambda}$
...	...

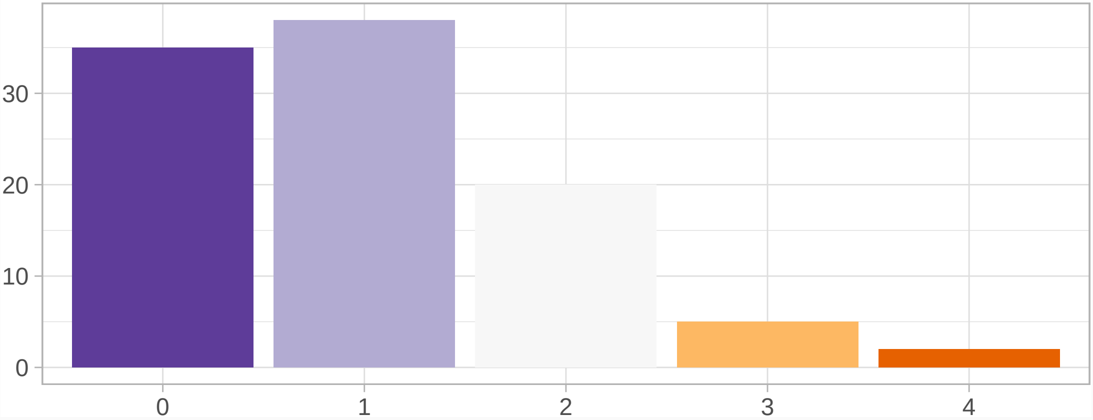
Poisson distribution

x	$P(X = x)$
0	$e^{-\lambda}$
1	$\lambda e^{-\lambda}$
...	...

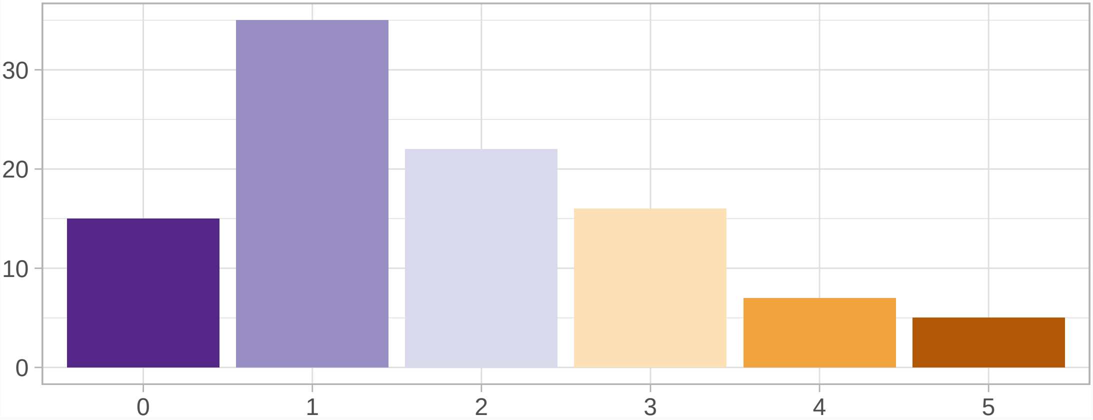
Fortunately, R has this pre-programmed

```
dpois(x=0,lambda=3) # equals exp(-3)
## [1] 0.04978707
```

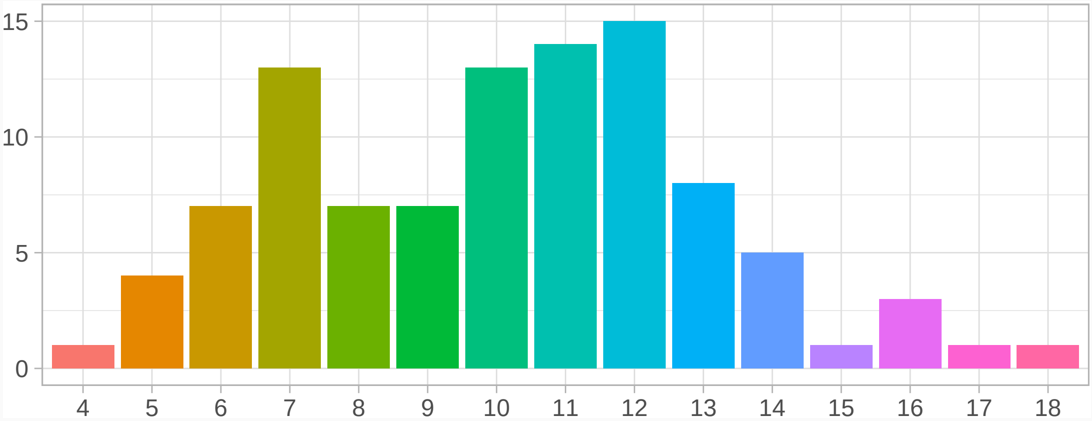
Hundred Poisson trials with $\lambda = 1$



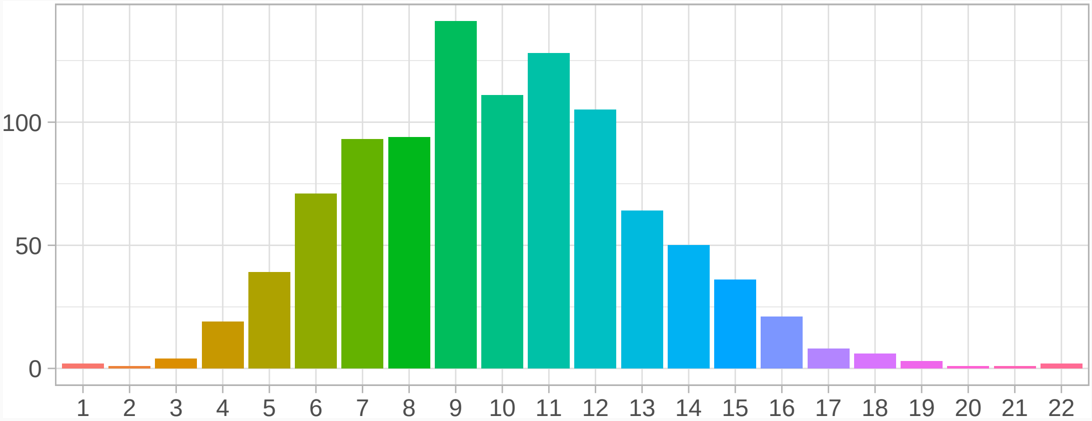
Hundred Poisson trials with $\lambda = 2$



Hundred Poisson trials with $\lambda = 10$



Thousand Poisson trials with $\lambda = 10$



Summary: Poisson distribution

- **notation**: $X \sim \text{Poisson}(\lambda)$
  - **range**: discrete, $x \geq 0$
  - **distribution**: $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$
  - **parameters**: λ the rate or expected number per sample
- **mean**: λ
  - **variance**: λ
  - **in R**: `rpois`, `dpois`

Continuous distribution

Normal (Gaussian) distribution

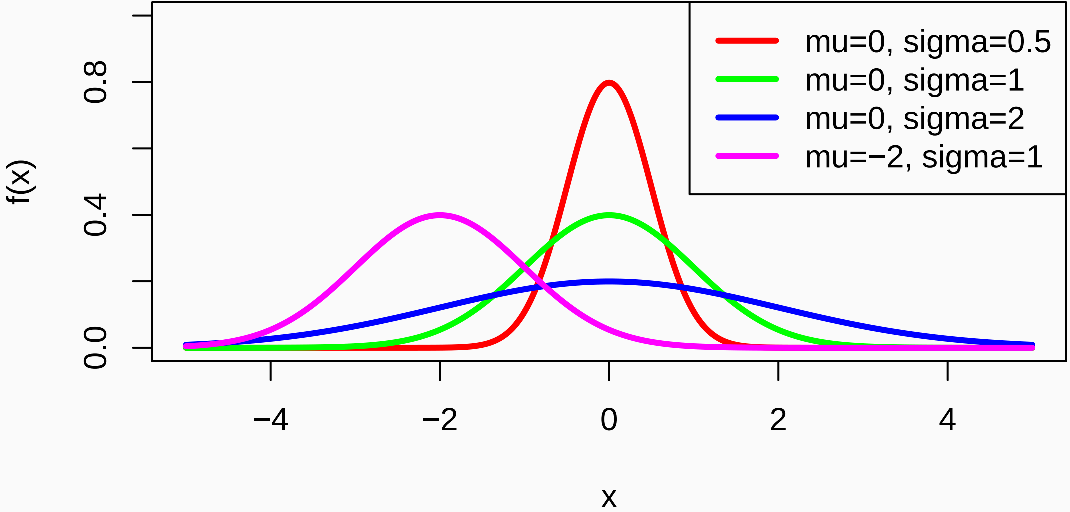
Context: Distribution of "adding lots of things together". Derived from *Central Limit Theorem*, which says that if you add a large number of independent samples from the same distribution the distribution of the sum will be approximately normal.

$X \sim \text{Normal}(\mu, \sigma^2)$ where μ is the mean and σ^2 the variance

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

Example: Practically everything.

Normal probability density function



Summary: Normal distribution

- **notation**: $X \sim \mathsf{N}(\mu, \sigma^2)$
  - **range**: continuous, all real values
- **distribution**: $f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
  - **parameters**: μ the mean and σ the standard deviation
- **mean**: μ
  - **variance**: σ^2
  - **in R**: `rnorm`, `dnorm`

Why do we love the Normal distribution

- If has nice properties, such as: if $X \sim \mathsf{N}(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim \mathsf{N}(0, 1)$
- It is a limiting distribution (*Central Limit Theorem*)
- It can be a good approximation for other distributions

Example: Approximating Binomial by Normal (1)

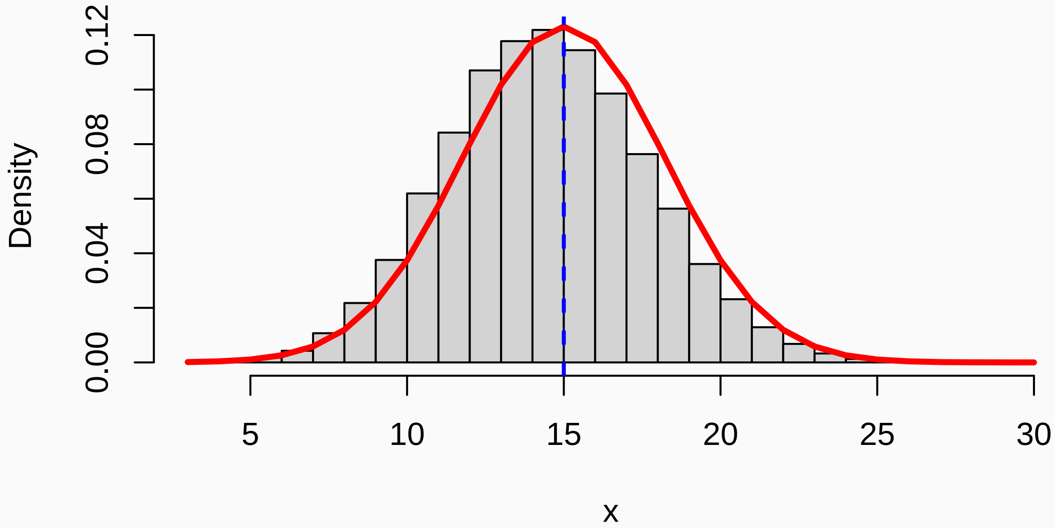
$X \sim \text{Bin}(N = 50, p = 0.3)$

Mean is $Np = 50 \times 0.3 = 15$

Variance is $Np(1 - p) = 50 \times 0.3 \times 0.7 = 10.5$

Therefore, X can be approximated by $Y \sim \text{N}(15, \sigma = \sqrt{10.5})$

Example: Approximating Binomial by Normal (2)



Conclusions about distributions

Common Distributions - Discrete

- When we have something that is dichotomous (either 0 or 1, negative/positive, false/true, male/female, present/absent):

Binomial(number of trials, probability)

- When we have something that is a discrete count, with no theoretical maximum, but with a common average:

Poisson(lambda)

Common Distributions - Discrete

- When we are recording the number of *failures* before a number of *successes*, or when we have something that is a discrete count with no theoretical maximum, and with more variation than Poisson:

$$\text{NegativeBinomial}(\text{number of successes, probability of success})$$

$$\text{NegativeBinomial}(\text{mean, overdispersion})$$

Common Distributions - Continuous

- When we have something that is continuous, symmetrical about the mean and unbounded:

Normal(mean, standard deviation)

- When we have something that is continuous, not symmetrical, and bounded at zero:

Exponential(rate)

Gamma(shape, rate)

Common Distributions - Continuous

- When we have something that is continuous, not symmetrical, and bounded at zero:

$$\text{Lognormal(logmean, logstdev)}$$

- When we have something that is continuous, and bounded between 0 and 1:

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

- Simple bounded distribution:

$$\text{Uniform(min, max)}$$

More? Check out in R:

```
?Distributions
```

Likelihoods

Fitting distributions to data

- So far, when talking about probability distributions, we assumed that we knew the parameter values
- And we wanted to know what data we might get from these distributions
- In the real world, it is usually the other way around
- A more relevant question might be:
  We have observed 3 births by a female during her 10 breeding attempts. What does this tell us about the true probability of getting a successful breeding attempt from this female? For the population?

Fitting distributions to data

We don't know what the probability of a birth is, but we can see what the probability of getting our data would be for different values:

```r
dbinom(x = 3, size = 10, prob = 0.1)
## [1] 0.05739563
```

Fitting distributions to data

We don't know what the probability of a birth is, but we can see what the probability of getting our data would be for different values:

```
dbinom(x=3,size=10,prob=0.9)
## [1] 8.748e-06
```

Fitting distributions to data

We don't know what the probability of a birth is, but we can see what the probability of getting our data would be for different values:

```
dbinom(x=3,size=10,prob=0.25)
## [1] 0.2502823
```

So we would be more likely to observe 3 births if the probability is 0.25 than 0.1 or 0.9

The likelihood

- This reasoning is so common in statistics that it has a special name:
- The likelihood is the probability of observing the data under a certain model
- The data are known, we usually consider the likelihood as a function of the model parameters $\theta_1, \theta_2, \ldots, \theta_p$

$$L = P(\theta_1, \theta_2, \ldots, \theta_p \mid \text{data})$$

- This is a very important concept

Likelihood functions

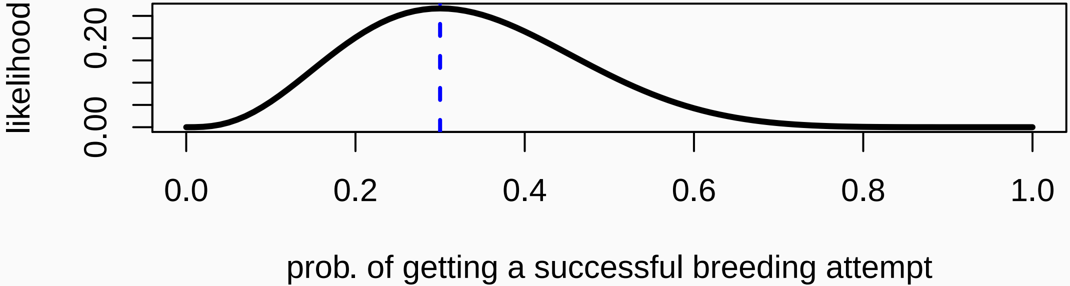
We may create a function to calculate a likelihood e.g.:

```r
lik.fun <- function(parameter){
  ll <- dbinom(x=3, size=10, prob=parameter)
  return(ll)
}

lik.fun(0.3)
## [1] 0.2668279

lik.fun(0.6)
## [1] 0.04246733
```

Maximize the likelihood (3 successes ot of 10 attempts)



The *maximum* of the likelihood is at value 0.3

The Maximum Likelihood

- There is always a set of parameters that gives you the highest likelihood of observing the data: the Maximum Likelihood Estimate(s) [MLEs]
- This can be calculated using:
- Trial and error (not efficient!)
- Compute the maximum of a function by hand (rarely doable in practice)
- An iterative optimization algorithm: `?optimize` (1 parameter) and `?optim` (> 1 parameter) in `R`

By hand: compute MLE of p from $Y \sim \text{Bin}(N = 10, p)$ with $k = 3$ successes

$P(Y = k) = \binom{k}{N} p^k (1-p)^{N-k} = L(p)$

$\log(L(p)) = \text{cte} + k\log(p) + (N-k)\log(1-p)$

We are searching for the maximum of L, or equivalently that of $\log(L)$

Compute derivate w.r.t. p: $\dfrac{d\log(L)}{dp} = \dfrac{k}{p} - \dfrac{(N-k)}{(1-p)}$

Then solve $\dfrac{d\log(L)}{dp} = 0$; the MLE is $\hat{p} = \dfrac{k}{N} = \dfrac{3}{10} = 0.3$

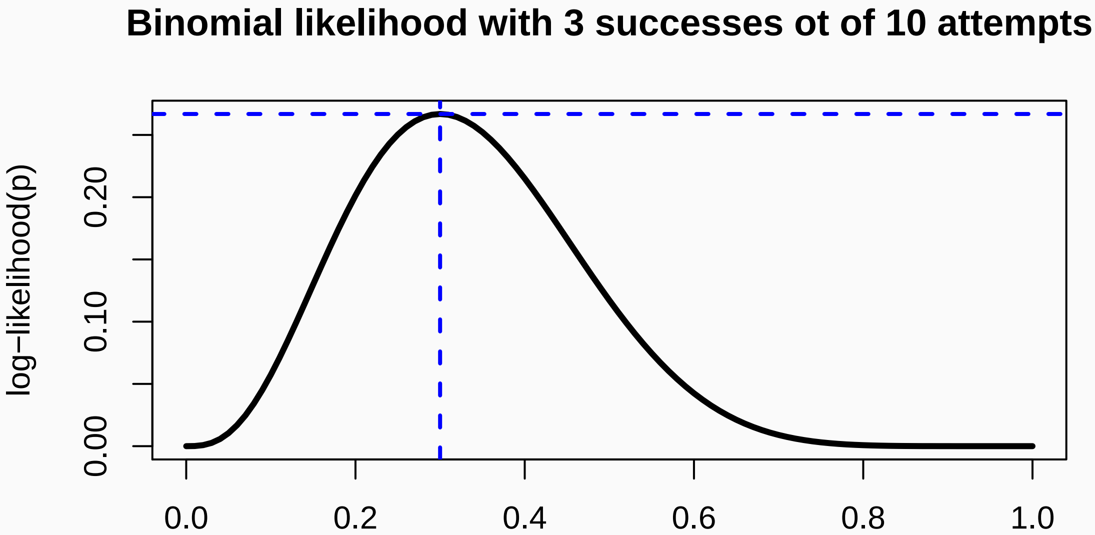
Here, the MLE is the proportion of observed successes

Using a computer: MLE of p from $Y \sim$ Bin$(N = 10, p)$ with $k = 3$ successes

```r
lik.fun <- function(parameter) dbinom(x=3, size=10, prob=parameter)
# ?optimize
optimize(lik.fun,c(0,1),maximum=TRUE)
## $maximum
## [1] 0.3000157
##
## $objective
## [1] 0.2668279
```

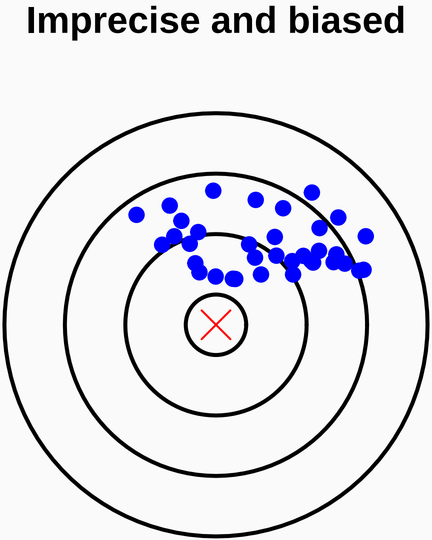
Use `optim` when the number of parameters is > 1.

Using a computer: MLE of p from $Y \sim \text{Bin}(N = 10, p)$ with $k = 3$ successes



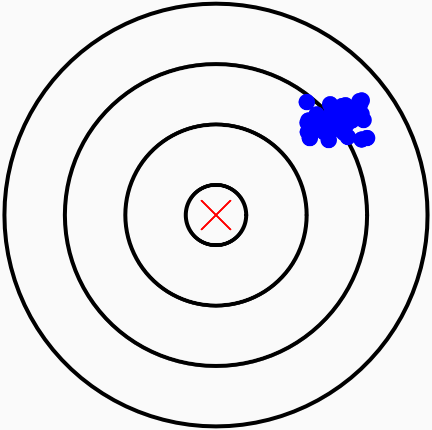
The Maximum Likelihood Estimate (MLE)

- The MLE is the best guess set of parameter values for our given data

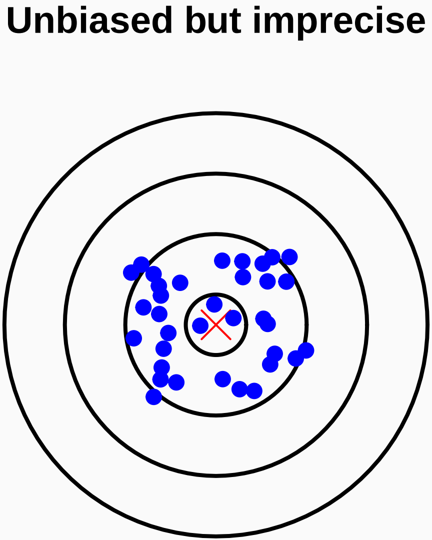

Imprecise and biased

A dart target, with the red cross representing the true parameter value

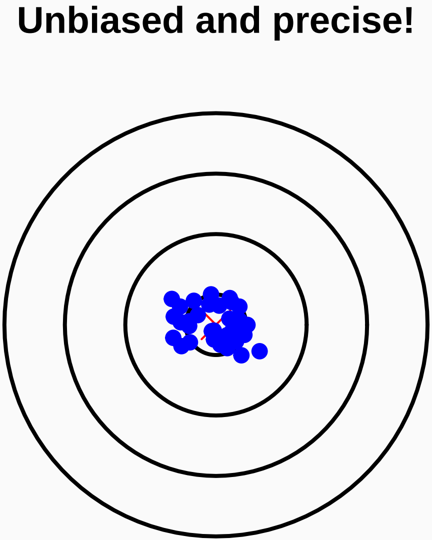
Precise but biased



A dart target, with the red cross representing the true parameter value



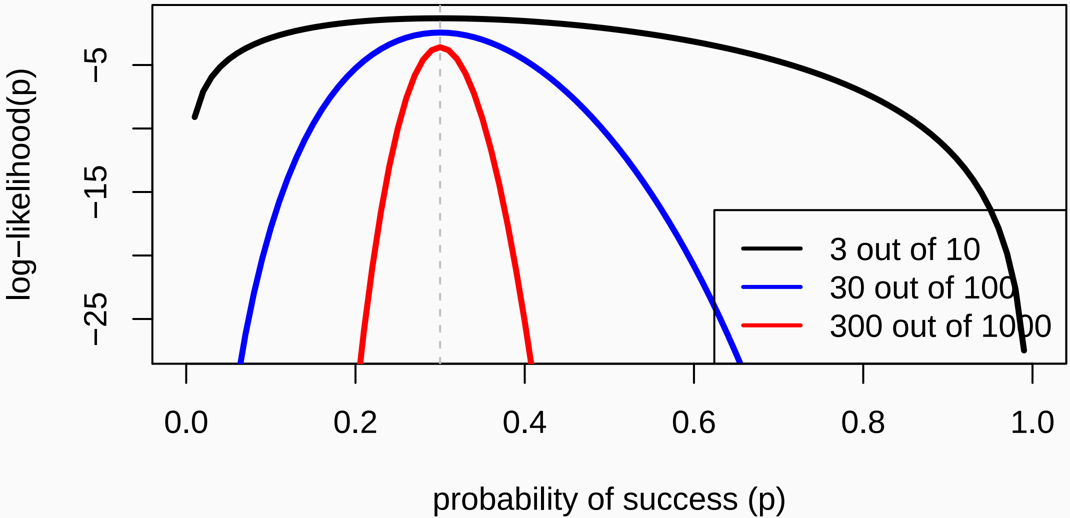
A dart target, with the red cross representing the true parameter value



The Maximum Likelihood Estimate (MLE)

- The MLE is the best guess set of parameter values for our given data
- But the chances of the true parameter values being close to the MLE is dependent on the amount of information in the data!

Binomial likelihood with increasing sample size



Confidence intervals: A refresher

Let's approach confidence intervals through simulations

Imagine you are measuring the temperature of a cup of water 10 times but you have an old really bad thermometer. The true temperature is 3 degrees Celsius and the standard deviation on the sampling error is 5.

```
# Simulate data:
mu <- 3
sigma <- 5
n <- 10
y <- rnorm(n = n, mean = mu, sd = sigma)
y
```

```
##  [1]  5.9276441  6.5473301  2.4534834  0.7325141  6.0294373 -6.0897798
##  [7]  6.1504928  1.6190795  1.5792013 -1.5966100
```

Apply linear regression

We will estimate a mean temperature by fitting an intercept only linear regression model:

```
m <- lm(y~1)
broom::tidy(m)
## # A tibble: 1 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)     2.34      1.29      1.82   0.103

confint(m)
##                  2.5 %   97.5 %
## (Intercept) -0.5749909 5.245549
```