Review of Probability Distributions

Alberto Garfagnini

Università di Padova

AA 2021/2022 - Stat Lect. 2

Pairing and Ordering of Objects

Unique pairing of objects

- given n objects, how many possible ways of selecting unique pairs, without caring about ordering ?
- let's consider a matrix $n \times n$
- every element in the matrix, except the leading diagonal, is a paring
- since the two parts on each side of the diagonal are identical (order does not count), we have

$$n_{pairs} = (n^2 - n)/2 = n(n-1)/2$$

Unique ordering of objects

- given n objects, how many possible ways of ordering them ?
- we have n options to select the first element
- $n-1$ for the second, $n-2$ for the third, ...
- therefore it is

$$n(n-1)(n-2)\dots 2 \cdot 1 = n!$$

Combinations and Permutations

- in the english language the word *"combination"* is used loosely, without specifying if the order of the object is relevant
- examples:
  - when buying an ice cream, we select a *combination* of mint, chocolate and stracciatella. We do not care about the order of the three flavours on the cone
  - the *combination* of my bike locker is 4-3-6-9. In this case, the order of the numbers really matters!
- when we select k elements from a set of n objects
  - if the order of selection is NOT important, we have a combination
  - but if the order matters, we have a permutation

➜ a permutation is and ordered combination

Permutations - ORDER MATTERS

- there are two types of permutations

Repetition IS allowed

- given n objects, how many sequences of r elements ($r \leq n$) can be built ?

  Example: given n letters, how many words of r characters can be built with those letters ?
  - each object (character) has n different possibilities, therefore it is

$$n^r$$

Repetition is NOT allowed

- given n objects, we select r elements ($r \leq n$) from the set
- how many unique selections are possible ?
  - there are n ways to select the first, $n-1$ for the second, and $n-r+1$ for the r-th
  - we get:

$$n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!} = {}^nP_r$$

- this is called permutations, ${}^nP_r$. Note that ${}^nP_n = n!$

Combinations - ORDER IS NOT IMPORTANT

- there are two types of combinations

Repetition is NOT allowed

- we now select r objects, as in the previous case, but we are not concerned about the order
- the number of ways of selecting r object from a set of n without regard to the order of selection is called combinations, ${}^{n}C_r$

$$ {}^{n}C_r = \frac{{}^{n}P_r}{n!} = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} $$

- this is the binomial coefficient, also called *n choose r*

Repetition IS allowed

- finally, the number of ways of choosing r objects from a set of n with replacement and without caring about the order is

$$ \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!} $$

- this is sometimes called *n multichoose r*

Application: the Birthday Paradox

The Problem

- in a large room, full of people, how many of them do you have to ask before there is a 50% chance that any of two, ore more, share a common birthday ?

- assuming $n = 365$ birthday/year and equally probable, we consider r people and we combine them so that they do not share a common birthday

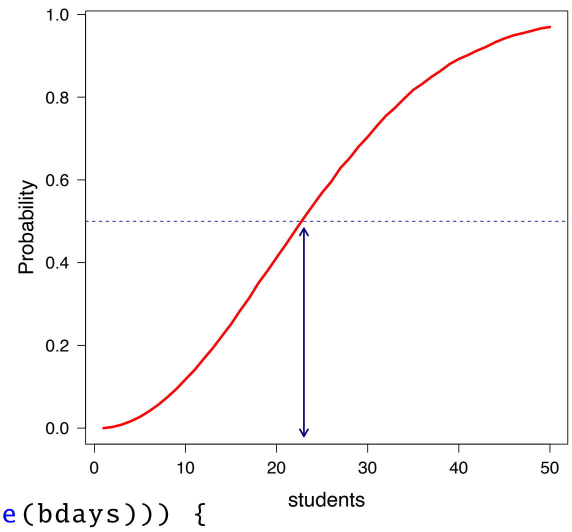
$$ A = n\,(n-1)\ \ldots\ (n-r+1) = \frac{n!}{(n-r)!} $$

- the way of assigning n birthday to r people is $B = n^r$
- the probability of no common birthday is A/B
- therefore the probability of at least one birthday is

$$ P(\text{birthday} \geq 1) = 1 - \frac{A}{B} = 1 - \frac{n!}{(n-r)!}\,\frac{1}{n^r} $$

Computation of the birthday problem

```
First element with prob>0.5:  23
```



R code

```
n_people_tot <- 50
pbday <- rep(0, n_people_tot)
for (k in 2:n_people_tot) {
  n_tests = 1E5; cb <- 0
  for (i in 1:n_tests) {
    bdays <- sample(1:365, k ,
                  replace=TRUE)
        if (length(bdays) > length(unique(bdays))) {
            cb = cb + 1
        }
    }
    pbday[k] <- cb/n_tests
    message(paste("k:", k, "pb(",k,"):",pbday[k]))
}
pfunc <- function(f, b) function(a) f(a,b)
p50_index <- Position(pfunc(`>`, 0.5), pbday)

message(paste("First␣element␣with␣prob>0.5:", p50_index))
```


R language note : closures

Anonymous functions

- can be used to create small function, not worth naming
- another important use is to create closures: functions written by functions

```
power <- function(exponent) function(x) x^exponent

square <- power(2)
square(2)
# [1] 4

cube <- power(3)
cube(2)
# [1] 8

pfunc <- function(f, b) {
              function(a) { f(a,b) }
}
p50 <- pfunc(`>`, 0.5)

x <- c(0.3, 0.51, 0.9)
p50(x)
# [1] FALSE  TRUE  TRUE
```

Probability Distributions

- two basic types: discrete distributions and continuous distributions
- discrete distribution : finite or countable set of possible outcomes of the random variable
- continuous distribution : a random variable can have outcomes in an interval of the real line
- probability densities are a way to specify probability distributions
- the cumulative distribution function (CDF) is defined by

$$F(x) = P(X \leq x)$$

- for discrete distributions:

$$F(x_j) = P(X \leq x_j) = \sum_j p_j$$

- while for continuous distributions:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

Probability Distributions

- with distribution functions, we compute the probability for intervals, $(c, d]$ as

$$P(c < X \leq d) = P(X \leq d) - P(X \leq c) = F(d) - F(c)$$

- the expectation, or expected value reflects the location of a distribution

$$E[X] = \sum_i x_i p(x_i) \qquad E[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

- the variance reflects the dispersion of the distribution:

$$var(X) = E[X - E[X]]^2 = E[X^2] - (E[X])^2$$

- properties:

$$\begin{array}{ll} E[a + bX] = a + bE[X] & var(a + bX) = b^2 var(X) \\ E[X + Y] = E[X] + E[Y] & var(X + Y) = var(X) + var(Y) + 2cov(X, Y) \end{array}$$

- with the covariance of the two variables

$$cov(X, Y) = E[(X - E[X])(Y - E[Y]] = E[XY] - E[X]E[Y]$$

Moments of a distribution

- they are analogous to the center-of-mass and to the momentum of inertia

Algebraic Moments

- the moment of order k about the origin is

$$\mu'_k \equiv E[x^k] = \int x^k f(x) dx \quad \text{and} \quad \sum_j x_j^k p_j$$

Central Moments

- the moment of order k about the mean are

$$\mu'_k \equiv E[(x-\mu)^k] = \int (x-\mu)^k f(x) dx \quad \text{and} \quad \sum_j (x_j-\mu)^k p_j$$

$$\begin{array}{ll} \mu'_0 = 1 & \mu_0 = 1 \\ \mu'_1 = \mu & \mu_1 = 0 \\ \mu'_2 = \mu + \sigma^2 & \mu_2 = \sigma^2 \end{array}$$

- the higher order moments become interesting only for studying the behavior of $f(x)$ for large $|x - \mu|$
- for a symmetric distribution, all odd central moments vanish ➜ non zero values are a possible measure of the skewness of a distribution

Probability Distributions in R

- all standard distributions available
- naming convention: a core name is associated with each distribution, and a prefix is appended to indicate the four basic associated functions:

- `d` for the probability density function (pdf)
- `p` for the cumulative density function (cdf)
- `q` for the quantile function
- `r` for the sampling from the distribution

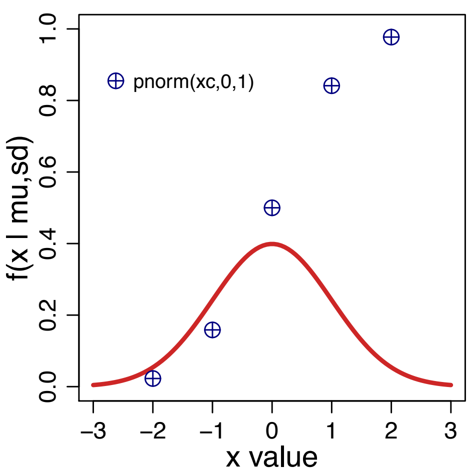
- note that `pcore_name()` and `qcore_name()` are one the inverse of one another

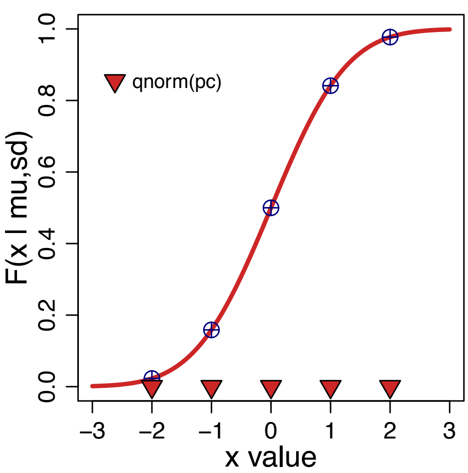
```
xc <- seq(-2,2,1)
pc <- pnorm(xc,0,1)
qc <- qnorm(pc)
```

xc: -2 -1 0 1 2

pc: 0.023 0.159 0.5 0.841 0.978

qc: -2 -1 0 1 2



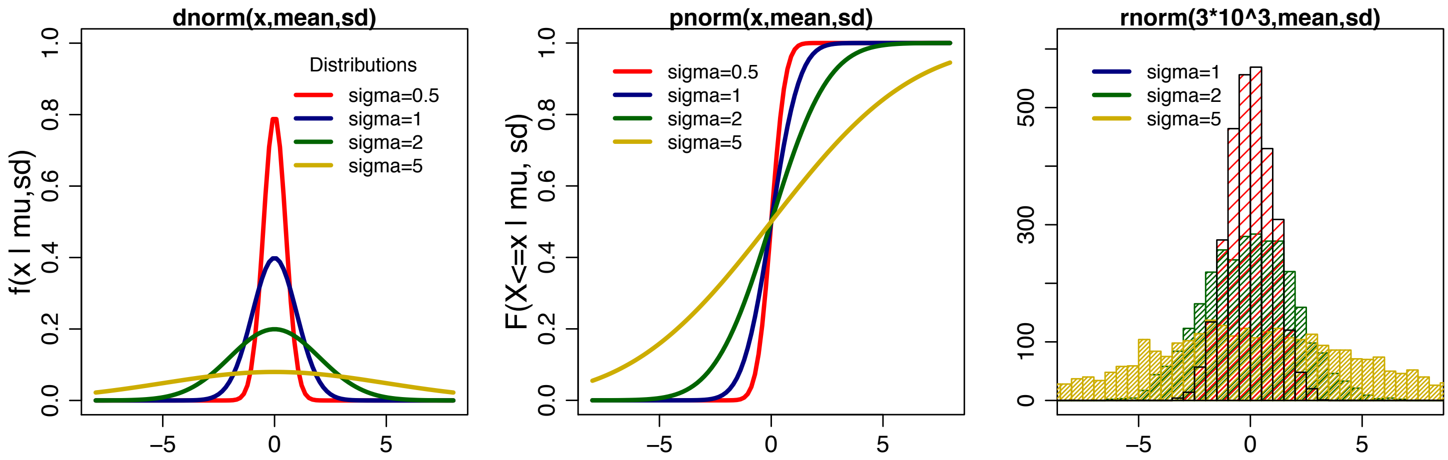


Standard Probability Distributions in R

Distribution	Core name	Parameters	Default values
Beta	`beta`	`shape1, shape2`	
Binomial	`binom`	`size, prob`	
Cauchy	`cauchy`	`location, scale`	0, 1
Chi-square	`chisq`	`df`	
Exponential	`exp`	`1/mean`	1
Fisher	`f`	`df1, df2`	
Gamma	`gamma`	`shape, 1/scale`	`NA`, 1
Geometric	`geom`	`prob`	
Hypergeometric	`hyper`	`m, n, k`	
Log-Normal	`lnorm`	`mean, sd`	0,1
Logistic	`logis`	`location, scale`	0,1
Normal	`norm`	`mean, sd`	0,1
Poisson	`pois`	`lambda`	
Student	`t`	`df`	
Uniform	`unif`	`min, max`	0,1
Weibull	`weibull`	`shape`	

Probability Distributions in R: normal distribution

- `dnorm(x, mean = 0, sd = 1)` gives a density of a normal distribution .i.e. the pdf
- `pnorm(q, mean = 0, sd = 1)` returns the distribution function, i.e. the cdf
- `rnorm(n, mean = 0, sd = 1)` generates random numbers from a normal distribution function
- `qnorm(p, mean = 0, sd = 1` is the quantile function



Standard Discrete Distributions

Bernoulli process

- it is a process with only two possible outcomes: success with probability p and failure with probability $1 - p$ (also called q, since $q = 1 - p$)
- if we call the two outcomes, 0 and 1, we can define $x \in [0, 1]$, and

$$P(X = 1) = p$$

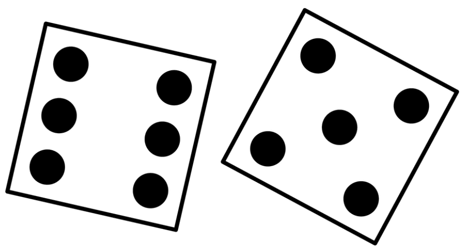
$$P(X = 0) = 1 - p = q$$

- the expected value and variance are

$$E[x] = p \quad \text{and} \quad Var(x) = p(1 - p)$$

Examples

- the toss of a coin
- the draw of a die



Binomial distribution

- the sum of n independent Bernoulli trials, follows a Binomial distribution

$$Bn(x \mid p, n) = \binom{n}{x} p^x (1-p)^{n-x}$$

- it gives the probability of $\mathtt{x}$ successes in $\mathtt{n}$ independent Bernoulli trials
- the expected value and variance are

$$E[x] = np \quad \text{and} \quad Var(x) = np(1-p)$$

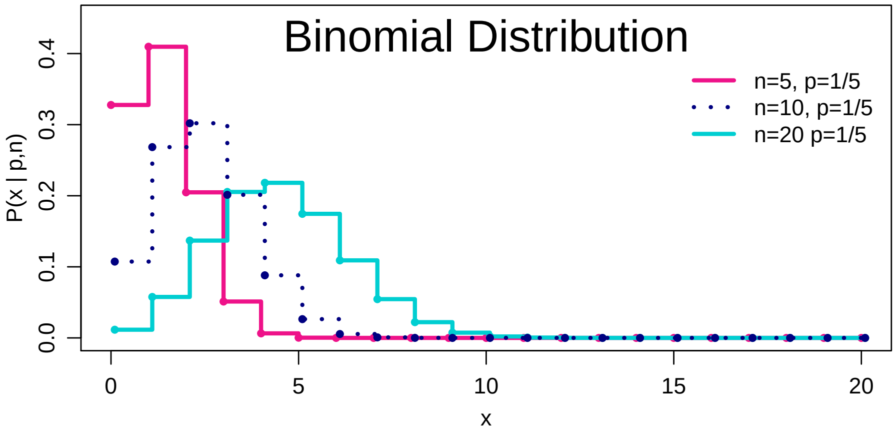
$$\sum_{j=0}^{n} \binom{n}{j} p^j \, (1-p)^{n-j} = (p + 1 - p)^n = 1$$

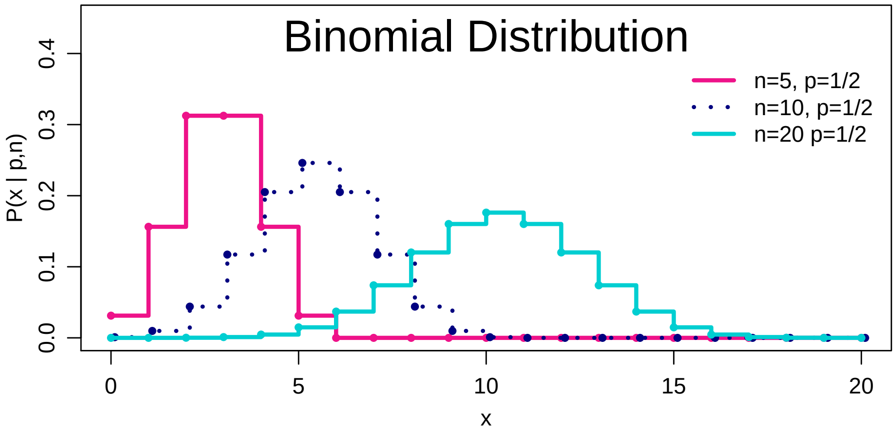
Examples

- multiple toss of a coin, or coins
- draw of dice
- drawing x red balls from an urn with n red and white balls (the fraction of red balls is p). Draws are done with replacement (➜ p remains constant)

Binomial distribution examples

- the distribution is symmetric when $p = 1/2$, and otherwise skew
- the distribution gets increasingly symmetric for higher values of n
- when n becomes large, it takes and approximate Gaussian shape





Binomial distribution - exercise

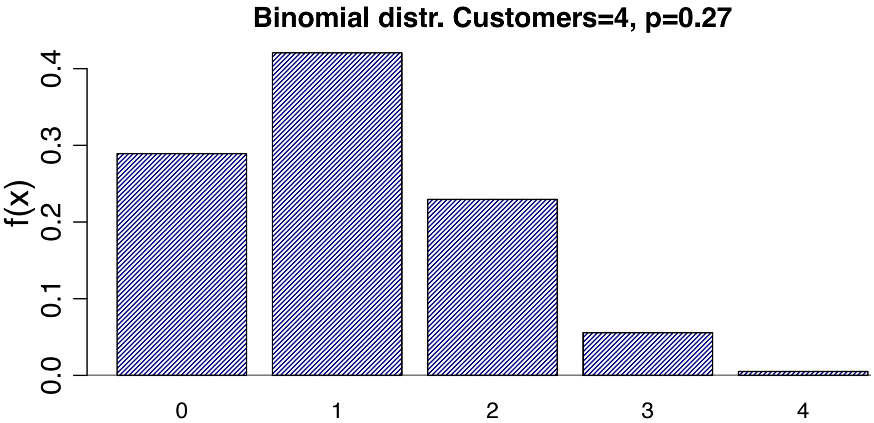
Problem

- in a restaurant 8 entrees of fish, 12 of beef and 10 of poultry are served
- what is the probability that 2 of the 4 next customers order fish entrees ?

Solution

```
cust <- 4
p <- 4/15
x <- 0:4
ap <- dbinom(x,cust,p)
barplot(ap, names=x, col='navy', xlab='x', ylab='f(x)', density=40,
        main = sprintf("Binomial distr. Customers=%d, p=%.2f",cust,p),
   cex.lab=1.5, cex.axis=1.25, cex.main=1.25, cex.sub=1.5)
cat(paste(c("P(2|np) = ", ap[3],'\n')))
```

```
P(2|np) = 0.229451851851852
```



Example: histogramming events

- we are interested in just the events contained in one bin of the histogram
  - A : we get the event of that particular bin (success)
  - $\overline{A}$: correspond to the events in any other bin (failure)
- the probability of having $x_\circ$ out of n events in the bin follows a Binomial distribution:

$$E[x] = np \quad \text{and} \quad Var(x) = np(1-p)$$

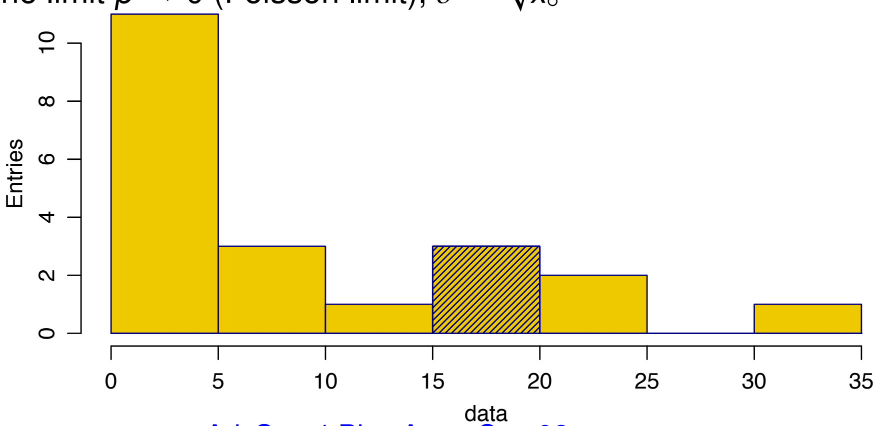
- p can be estimated as the ratio $p = x_\circ/n$:

$$E[x] = np = n\frac{x_\circ}{n} = x_\circ \quad \text{and} \quad Var(x) = x_\circ(1 - \frac{x_\circ}{n})$$

➜ the error on the number of the events is not $\sqrt{x_\circ}$, but a smaller quantity, $\sqrt{x_\circ(1-p)}$. Only in the limit $p \to 0$ (Poisson limit), $\sigma = \sqrt{x_\circ}$

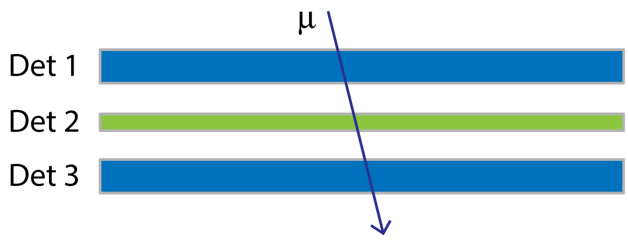
$\sqrt{3} = 1.73$

$\sqrt{3(1-3/21)} = 1.60$



Example: detection efficiency

- we want to compute the efficiency of a detector and evaluate the uncertainty on the measurement
- a muon-like signal has been registered by `Det1` and `Det3`
- what is the detection efficiency of our `Det2` ?



- detection is a Bernoulli process:

$$\epsilon_2 = \frac{N_{det2}}{N_{det1\&det3}} \quad with \quad N_{det2} \subset N_{det1\&det3}$$

- since we are interested in a relative number of success in a trial,

$$E\left[\frac{r}{n}\right] = \frac{1}{n}E[r] = p \quad \text{and} \quad Var\left(\frac{r}{n}\right) = \frac{1}{n^2}V(r) = \frac{p(1-p)}{n} = \frac{pq}{n}$$

- in our case, p is the ratio of events detected with `Det2` with respect to those seen by both `Det1` and `Det3`
- therefore:

$$\sigma(\epsilon_2) = \sqrt{\frac{\epsilon_2(1-\epsilon_2)}{N_{det1\&det3}}}$$

The drunk-man and the home keys problem

The background information

- a man comes back home pretty drunk
- he has 8 keys and tries them randomly to unlock his door apartment
- after each trial he loses memory
- we watch him and bet on the attempt on which he will succeed
- $n_{try} = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \ldots$
- on which number would you bet ?

The problem

- E_j : the door gets unlocked in attempt j, with $j = 1, 2, \ldots$
- we know that: $P(E_i | \overline{\bigcup_{j<i} E_j}) = 1/8$

  $f(1) = P(E_1) = p = 1/8$

  $f(2) = P(E_2 \cdot \overline{E}_1) = P(E_2 | \overline{E}_1) \cdot P(\overline{E}_1) = p \cdot (1-p)$

  $f(3) = P(E_3 \cdot \overline{E}_2 \cdot \overline{E}_1) = P(E_3 | \cdot \overline{E}_2 \cdot \overline{E}_1) \cdot P(\overline{E}_2 | \cdot \overline{E}_1) \cdot P(\cdot \overline{E}_1) = p \cdot (1-p)^2$

  $f(x) = p \cdot (1-p)^{x-1}$

Geometric distribution

- our probabilities follow a geometric distribution with $p = 1/8$

  $f(1) = p = 1/8 = 0.125$ ✓ our best bet!

  $f(2) = p(1-p) = 1/8(7/8) = 0.109$

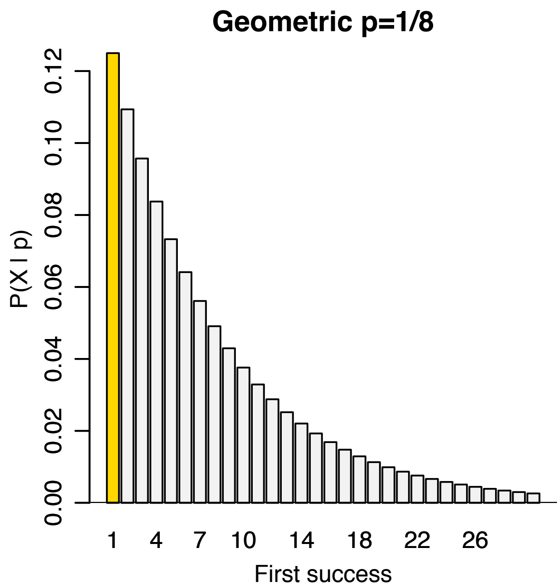
  $f(3) = p(1-p)^2 = 0.096$

  $f(4) = p(1-p)^3 = 0.084$

  ...

- the geometric distribution gives the number of trials to get the first success

$$Geo(x|p) = p\,(1-p)^x$$



- the expected value and variance are

$$E[x] = \frac{1}{p} \quad \text{and} \quad Var(x) = \frac{1-p}{p^2}$$

- useful relations:

$$P(x \le r) = 1 - (1-p)^r = q^r \quad \text{and} \quad P(x > r) = 1 - q^r$$

Geometric distribution examples (1)

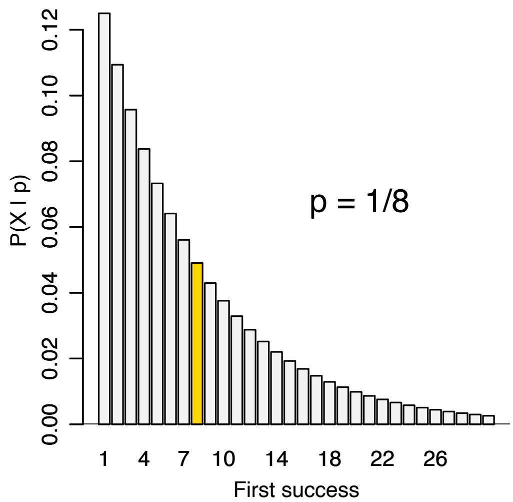
Drunk-man

- the first trial is the most probable
- but

$$E[X] = 1/p = 8$$

and

$$\sigma = \sqrt{(1-p)/p^2} = 7.5$$

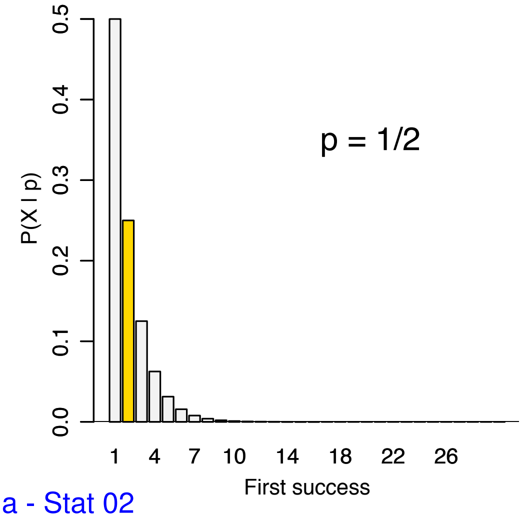


Coin tossing

- if we apply it to the tossing of one coin, we get
- $p_{max} = p = 1/2$
- and $E[X] = 1/p = 2$

  and

  $\sigma = \sqrt{(1-p)/p^2} = 1.4$



Geometric distribution examples (2)

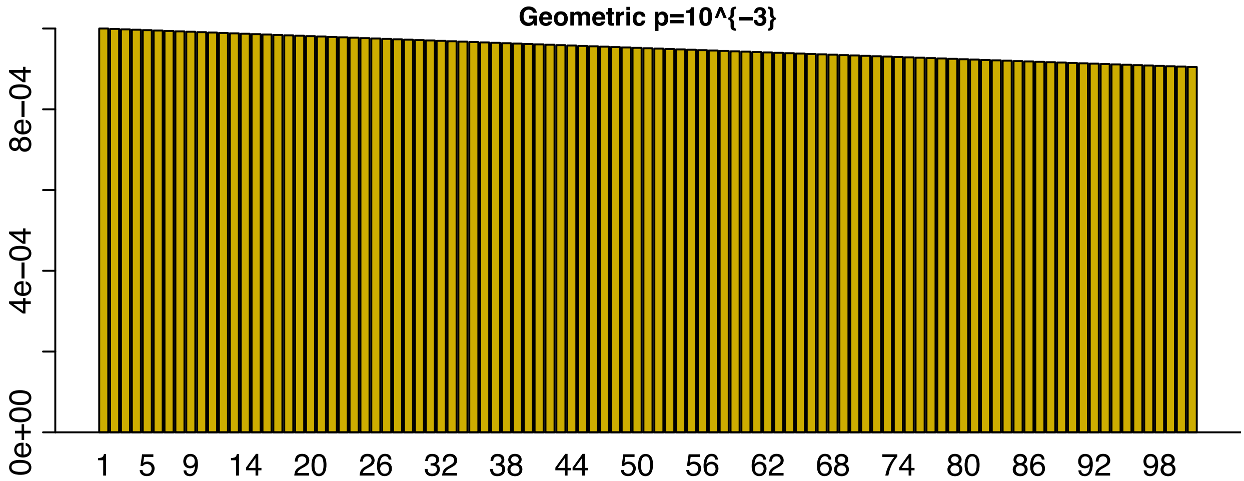
Rare Events

- let's decrease the probability of the event

$$E[X] = 1/p = 10^3$$

$$Var(X) = \frac{\sqrt{1-p}}{p} \xrightarrow[p\to 0]{} \frac{1}{p}$$

- rare moments might happen at any moment
  (even if they have a negligible probability to happen at any moment)



Geometric distribution in R

- given $x = \{1, 2, 3, \ldots\}$ as the number of trials for the first success
  an alternative representation uses
- $y = \{0, 1, 2, \ldots\}$ as the number of failures before the first sucess
- the two representations are equivalent:

```
y = x - 1

f(x) = p(1-p)^x = 1 - q^x
F(x) = 1 - (1-p)^x = 1 - q^x
E[x] = (1-p)/p     Var[x] = (1-p)/p^2

f(y) = p(1-p)^y
F(y) = 1 - (1-p)^(y+1)
E[y] = (1-p)/p     Var[x] = (1-p)/p^2
```

- the geometric distribution in R

```
Geometric                   package:stats                   R Documentation
The Geometric Distribution
Usage:

     dgeom(x, prob, log = FALSE)
     ...
Arguments:
    x, q: vector of quantiles representing the number of failures in a
          sequence of Bernoulli trials before success occurs.
```

Multinomial distribution

- it is a generalization of the binomial distribution to the case with more than 2 possible outcomes
- labeling the disjoint outcomes $A_1, A_2, \ldots, A_r$, we define $P(A_j) = p_j$, with $1 \leq j \leq r$
- in n independent trials, x_j denotes the number of times that A_j occurs
- assuming, by construction, $n = x_1 + x_2 + \ldots + x_r$, we have

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_r = x_r | p_1, p_2, \ldots p_r, n) = \frac{n!}{x_1! x_2! \ldots x_r!} p_1^{x_1} p_2^{x_2} \ldots p_r^{x_r}$$

Properties

- the expectation for class A_j is $E[x_j] = np_j$
- the variance for class A_j is $Var(x_j) = np_j(1 - p_j)$
- the covariance for classes A_i, A_j is $cov(x_i, x_j) = -n\, p_i\, p_j$
- when n becomes large, the distribution tends to a multinormal distribution

Multinomial distribution - exercise

Problem

- in a certain town, at 20:00, 30% of the TV audience watches the news, 25% a TV show, and the rest other programs
- What is the probability that, selecting 7 random viewers, exactly 3 watch the news and at least 2 watch the TV show ?

Solution

- the probabilities are $p_1 = 3/10$, $p_2 = 1/4$, $p_3 = 9/20$
- the sum of the trials $i + j + k = 7$
  - we write

$$P\left(i, j, k \middle| n = 7\right) = \frac{7!}{i!\; j!\; k!} \left(\frac{3}{10}\right)^i \left(\frac{1}{4}\right)^j \left(\frac{9}{20}\right)^k$$

  - and we compute

$$\begin{aligned} P(i = 3, j \geq 2 | n = 7) &= P(3, 2, 2|7) + P(3, 3, 1|7) + P(3, 4, 0|7) \\ &= \tfrac{7!}{3!\; 2!\; 2!} \left(\tfrac{3}{10}\right)^3 \left(\tfrac{1}{4}\right)^2 \left(\tfrac{9}{20}\right)^2 + \tfrac{7!}{3!\; 3!} \left(\tfrac{3}{10}\right)^3 \left(\tfrac{1}{4}\right)^3 \left(\tfrac{9}{20}\right) \\ &+ \tfrac{7!}{3!\; 4!} \left(\tfrac{3}{10}\right)^3 \left(\tfrac{1}{4}\right)^4 \simeq 0.103 \end{aligned}$$

Multinomial distribution marginalization

- let suppose we have a multinomial distribution $P(X_1, X_2, \ldots X_r)$ and we want to find the marginal probability $P(X_1)$

$$
\begin{aligned}
P(X_1) &= \sum_{x_2+x_3+\ldots+x_r=n-x_1} \frac{n!}{x_1!x_2!\ldots x_r!} p_1^{x_1} p_2^{x_2} \ldots p_r^{x_r} \\
&= \frac{n!}{x_1!(n-x_1)!} p_1^{x_1} \sum_{x_2+x_3+\ldots+x_r=n-x_1} \frac{(n-x_1)!}{x_2!\ldots x_r!} p_2^{x_2} \ldots p_r^{x_r} \\
&= \frac{n!}{x_1!(n-x_1)!} p_1^{x_1} (p_2+\ldots+p_r)^{n-x_1} \\
&= \frac{n!}{x_1!(n-x_1)!} p_1^{x_1} (1-p_1)^{n-x_1}
\end{aligned}
$$

- where the multinomial expansion has been used, and also the fact that $p_1 + p_2 + \ldots + p_r = 1$
- the obtained distribution coincides with the binomial distribution

Poisson process

- let's consider an event that might happen at a given time, with the following conditions:
- the probability of 1 count in Δt is proportional to Δt itself, with Δt a 'small' value
- calling r, the intensity of the process,

$$p = P('1 \text{ count in } \Delta t') = r\Delta t$$

- moreover:
- $P(\geq 2 \text{ counts}) \lll P(1 \text{ count})$
- what happens in one interval does not depend on other intervals ➜ it has a memory-less property

Examples

- accidents occurring at an intersection
- γ-s emitted from a radioactive substance
- customers entering a post office
- earthquakes in Italy

Poisson distribution

- the Poisson distribution can be derived by the Binomial distribution, in the limit where the rate of success, p, is very small
- we divide a finite time interval, T, in n small intervals:

$$T = n\,\Delta T$$

- and we consider the possible occurrence of an event, an independent Bernoulli trial, in each small interval Δt

$$p = r\,\Delta T = r\frac{T}{n}$$

- if the number of of trials is large, the total number of successes, np, is however considerable: $np = rT = \lambda$
- mathematically, in the limit $p \to 0$, $n \to \infty$ and $np = \lambda$ remaining constant, we get

$$\mathrm{Bn}\left(r\middle|n\,p\right) \to Poi\left(r\middle|\lambda\right)$$

- λ depends only on the intensity of the process, r, and on the finite time of observation

$$\mathrm{Poi}(r|\lambda) = \frac{\lambda^r}{r!}\exp(-\lambda)$$

Poisson distribution

- Given the Poisson distribution function:

$$\mathrm{Poi}\left(r\middle|\lambda\right) = \frac{\lambda^r}{r!}\exp(-\lambda)$$

- the expected value and variance are

$$E[x] = \lambda \quad \text{and} \quad Var(x) = \lambda$$

- Asymptotically, for growing λ values, the Poisson distribution becomes identical to the normal distribution

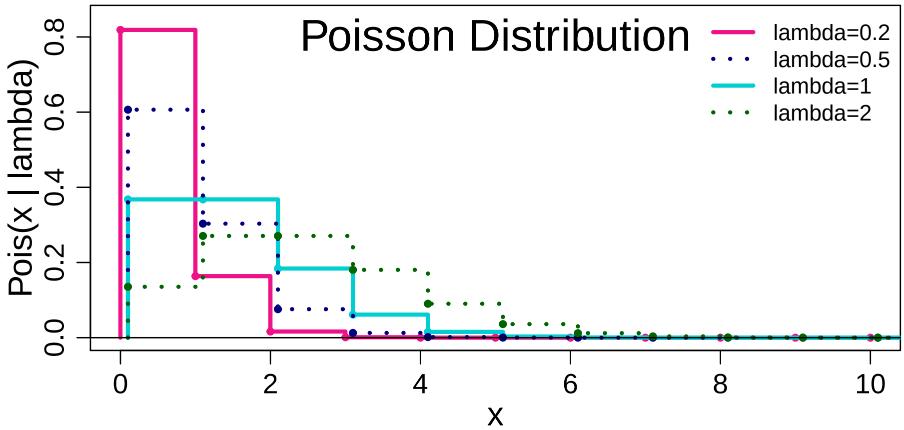
  the similarity is rather close already at $\lambda = 20$
- an interesting property is:

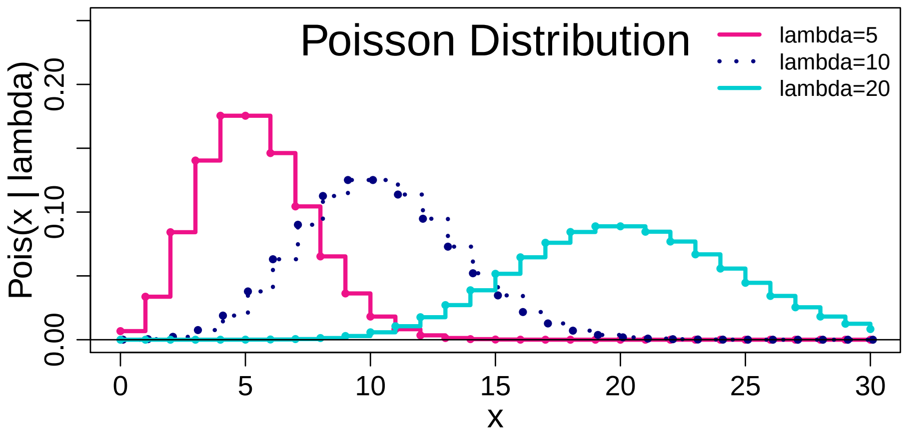
$$\mathrm{Poi}\left(r\middle|\lambda\right) = Poi\left(r-1\middle|\lambda\right)\frac{\lambda}{r}$$

- it is possible to demonstrate that the sum of any independent Poisson variables is itself a Poisson variable with mean value equal to the sum of the individual means

Poisson distribution examples

- the distribution is very asymmetric for small λ and it has a tail to the right of the mean
- the distribution gets increasingly symmetric for higher values of λ
- already for $\lambda = 20$ is very similar to the normal distribution (but it has only integer values)

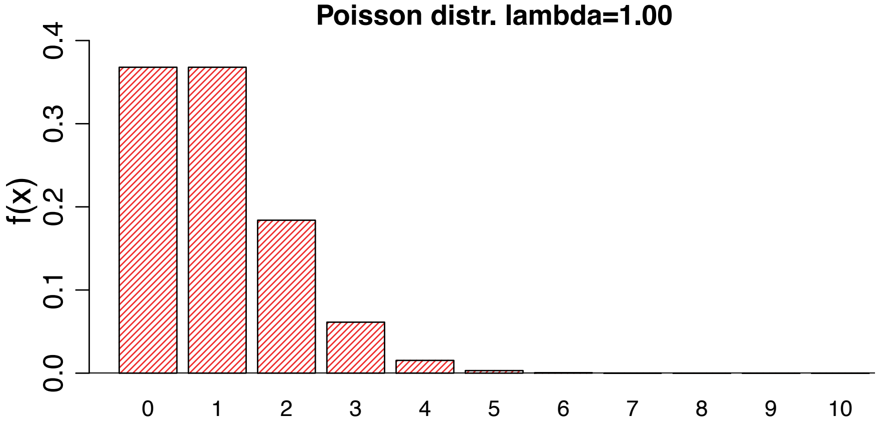




Poisson distribution - exercise 1

Problem

- the average number of received wrong phone calls per week is 7
- what is the probability to get, tomorrow, A) two wrong calls ? B) at least one wrong call ?



Solution

- assuming we get a large number of calls, the number of wrong calls follows, to a good approximation, a Poisson distribution
- we assume $\lambda = 1$

```
lambda <- 1
x <- 0:10
ap <- dpois(x,lambda)
barplot(ap, names=x, col='firebrick2', xlab='x', ylab='f(x)', density=30,
        main = sprintf("Poisson␣distr.␣lambda=%.2f",lambda),
        ylim=c(0,0.415),
        cex.lab=1.5, cex.axis=1.25, cex.main=1.25, cex.sub=1.5)
cat(paste(c("P(2|lambda)␣=␣", ap[3],'\n')))
cat(paste(c("P(>=1|lambda)␣=␣", 1 - ap[1],'\n')))
```

```
P(2|lambda) = 0.184
P(>=1|lambda) = 0.632
```

Poisson distribution - exercise 2

Problem

- a radioactive substance emits on average 3.9 α/s per gram
- compute the probability that, in the next second, the number of emitted alpha particles is

  A) at most 6

  B) at least 2

  C) at least 3 and at most 6

Solution

- every gram of element has n atoms
- From the information we have, $E[X] = np = \lambda = 3.9$

$$P(x|\lambda) = \frac{\lambda^x}{x!} \exp(-\lambda)$$

$$\text{A)} \quad P(x \leq 6) = \sum_{x=0}^{6} \frac{3.9^x}{x!} \exp(-3.9)$$

$$\text{B)} \quad P(x \geq 2) = 1 - P(x \leq 1) = 1 - \sum_{x=0}^{1} \frac{3.9^x}{x!} \exp(-3.9)$$

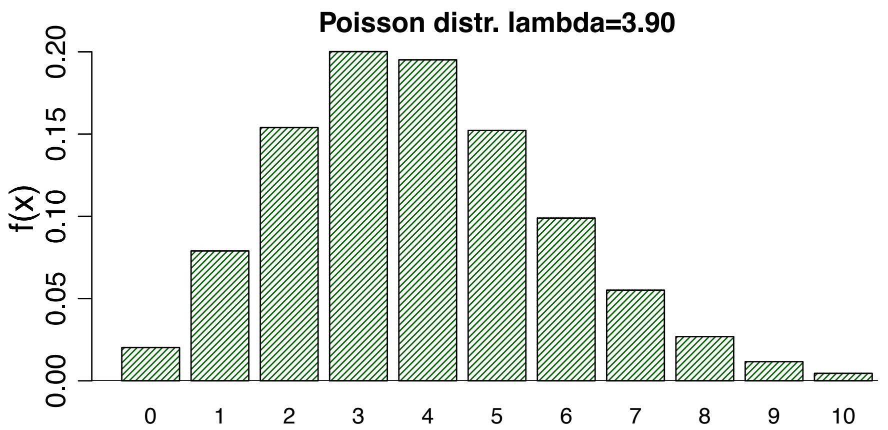
$$\text{C)} \quad P(3 \leq x \leq 6) = \sum_{x=3}^{6} \frac{3.9^x}{x!} \exp(-3.9)$$

Poisson distribution - exercise 2

```
P(<=6) = 0.899483035093612

P(>=2) = 0.900814633915558

P(2<X<=6) = 0.646357932463829
```




```
lambda <- 3.9
x <- 0:10
ap <- dpois(x,lambda)

barplot(ap, names=x, col='darkgreen', xlab='x', ylab='f(x)', density=30,
        main = sprintf("Poisson distr. lambda=%.2f",lambda),
        ylim=c(0,0.21),
        cex.lab=1.5, cex.axis=1.25, cex.main=1.25, cex.sub=1.5)
abline(0,0)

P_6 = sum(ap[x<=6])
P_2 = 1 - sum(ap[x<=1])

cat(paste(c("P(<=6) = ", P_6,'\n')))
cat(paste(c("P(>=2) = ", P_2,'\n')))

pp <- ppois(x, lambda)
P_36 = pp[x==6] - pp[x==2]
cat(paste(c("P(2<X<=6) = ", P_36,'\n')))
```

Pascal or Negative Binomial distribution

- the probability of obtaining the r-th success in n trials, is given by the Negative Binomial, or Pascal, distribution
- since in $n-1$ trials we had $r-1$ successes, the probability is given by the Binomial distribution:

$$\mathrm{Bn}\left(r \middle| n, p\right) = \binom{n-1}{r-1} p^{r-1} (1-p)^{n-1-r+1} = \binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$$

- but we got the r-th success at the n-th trial, therefore

$$\mathrm{Bneg}\left(r \middle| n, p\right) = \binom{n-1}{r-1} p^{r} (1-p)^{n-r}$$

- the expected value and variance are

$$E[x] = \frac{r}{p} \quad \text{and} \quad Var(x) = \frac{r(1-p)}{p^2}$$

Pascal distribution - exercise

Problem

- Ann and Maggie are playing cards until one of them wins 5 games
- suppose all games are independent and the probability that Ann wins is 58%

  A) what is the probability that they complete in 7 games

  B) if the series ends in 7 games, what is the probability that Ann wins ?

Solution to A

- `X`: number of games played until Ann wins 5 games
- `Y`: number of games played until Maggie wins 5 games
- both `X` and `Y` follow a Pascal distribution

$$P(X=7, r=5) = \binom{6}{4} 0.58^5 0.42^2 = 0.174$$

$$P(Y=7, r=5) = \binom{6}{4} 0.42^5 0.58^2 = 0.066$$

- we get $P(X=7, r=5) + P(Y=7, r=5) = 0.24$

Pascal distribution - exercise

Solution to B

- A: Ann wins
- B: the series ends in 7 games

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(X=7)}{P(X=7)+P(Y=7)} = \frac{0.17}{0.24} = 0.71$$

Solution with R

```
dnbinom(x, size, prob, mu)
```

```
The negative binomial distribution with 'size' = n and 'prob' = p
...
for x = 0, 1, 2, ..., n > 0 and 0 < p <= 1.

This represents the number of failures which occur in a sequence
of Bernoulli trials before a target number of successes is
reached.  The mean is mu = n(1-p)/p and variance n(1-p)/p^2.

P_Ann    <- dnbinom(2,5,0.58) # 0.173672
P_Maggie <- dnbinom(2,5,0.42) # 0.0659468
```


Exercise : Binomianl/Poisson

Defective screws

- a company produces screws
- the probability of a screw to be defective is $p = 0.015$
- a box with $n = 100$ screws is packaged.

  Compute:

A) the probability that all screws are non defective

B) the defective screws distribution comparing the Binomial and Poisson distributions

C) how many extra screws should the box contain in order to have $n = 100$ non defective screws with probability greater than 80%