Chapter 4: Univariate Probability Distributions

Wang Shujia

Contents

1 Discrete Univariate Distributions **2**
1.1 Bernoulli and Binomial Distributions 2
1.2 Geometric and Negative Binomial Distribution 4
1.3 Poisson Distribution 5

2 Continuous Univariate Distributions **8**
2.1 Uniform Distribution and Normal Distribution 9
2.2 Exponential Distribution and Gamma Distribution 12
2.3 Beta, Cauchy, Lognormal and Weibull 15

1 Discrete Univariate Distributions

Objectives

You will learn in this section

How to choose a reasonable discrete model under a variety of physical circumstances

Common discrete distributions

- Discrete uniform distribution
- Binomial
- Geometric
- Negative binomial
- Poisson
- Hypergeometric Distribution

1.1 Bernoulli and Binomial Distributions

Bernoulli trial

The binomial distribution is based on a *Bernoulli trial*, which is a random experiment in which there are only two possible outcomes: success (S) and failure (F).

We conduct the Bernoulli trial and let

$$X = \begin{cases} 1 & \text{if the outcome is } S \\ 0 & \text{if the outcome is } F \end{cases}$$

Then X is called a **Bernoulli random variable**.

If the probability of success is p then the probability of failure must be $1 - p = q$, then the pmf of X is

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

It is easy to calculate $\mu = \mathrm{E}X = p,\ \sigma^2 = pq$, and $M(t) = pe^t + q$.

Binomial Experiment

The binomial experiment has three defining properties:

- Bernoulli trials are conducted n times,
- the trials are independent,
- the probability of success p does not change between trials.

Binomial Distribution

If X is the *number of successes* in n Bernoulli trials, then X is a **Binomial Random Variable**, and is denoted by

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

The pmf is:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} (x = 0, 1, \cdots, n)$$

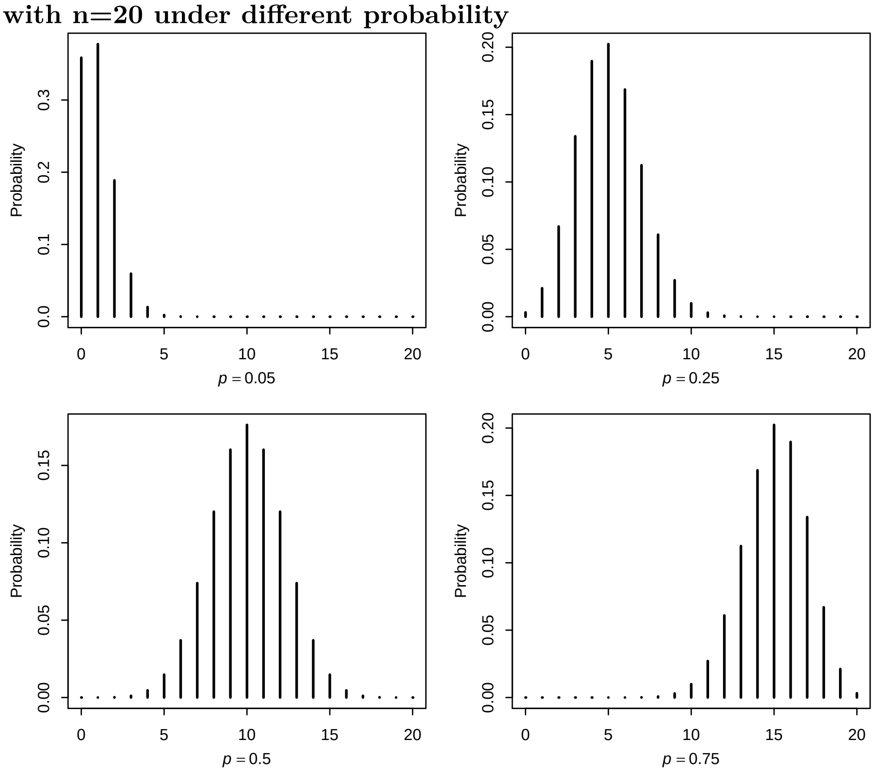
Mean $\mu = np$, variance $\sigma^2 = np(1-p)$, mgf $M(t) = (pe^t + q)^n$.

Remark: the combination $\binom{n}{x}$ can be computed by R function: `choose(`n, x`)`.

R functions for binomial

- `dbinom(z,n,p)`: calculate $P(X = z)$.
  - "d" stands for *distribution*.
- `pbinom(z,n,p)`: calculate$P(X \leq z)$.
  - "p" stands for *probability*.
- `rbinom(1000,n,p)`: generate 1000 samples from X.
  - "r" stands for *random numbers*.

Binomial with n=20 under different probability



Binomial Distributions Tends to Normal

If

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

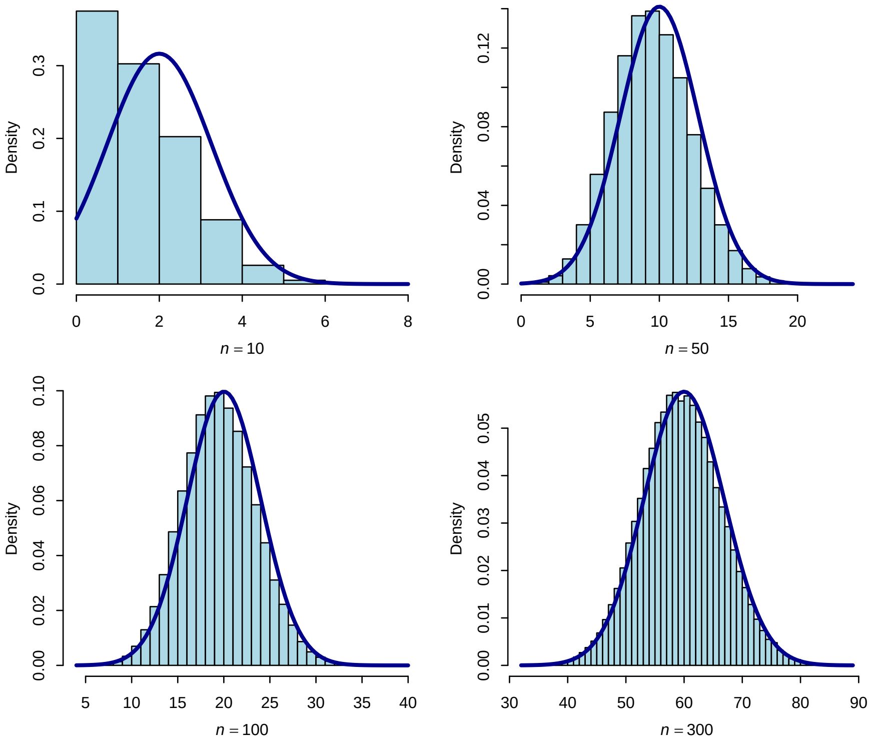
then

$$\frac{X-np}{\sqrt{npq}} \xrightarrow{d} \text{N}(0,1)$$

i.e. as $n \to \infty$,

$$X \approx \text{N}(np, npq)$$

Bin(n,p) with p=0.2 and n=10,50,100,300



1.2 Geometric and Negative Binomial Distribution

Geometric Distribution

Suppose that we conduct Bernoulli trials repeatedly, noting the successes and failures. Let X be the *number of trials until the first success.* Then X has pmf

$$f(x) = P(X = x) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots$$

We say that X has a **Geometric distribution** and we write $X \sim \text{geom}(p)$.
Mean $\mu = 1/p$, variance $\sigma^2 = q/p^2$ and $M(t) = p/(1-qe^t)$.
The R functions are dgeom(x, prob), pgeom(), qgeom() and rgeom()

Negative Binomial Distribution

Suppose that we conduct Bernoulli trials repeatedly, let X count the *number of failures before r successes*. Then X has pmf

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \ldots$$

We say that X has a **Negative Binomial** distribution and write $X \sim \text{NB}(r, p)$.

The associated R functions are dnbinom(x, size, prob) (“d” may be “p”,”q” and “r”).

Mean $\mu = rq/p$, variance $\sigma^2 = rq/p^2$ and $M(t) = p^r(1-qe^t)^{-r}$.

Useful Relationships

1. If n independent random variables, $X_1, X_2, \ldots, X_n \sim \text{Geom}(p)$, then $\sum X_i \sim \text{NB}(n, p)$.
2. If n independent random variables, $X_i \sim \text{NB}(r_i, p)(i = 1, 2, \ldots n)$, then $\sum X_i \sim \text{NB}(\sum r_i, p)$.
3. When $n = 1$, $\text{NB}(1, p) = \text{Geom}(p)$.

1.3 Poisson Distribution

Poisson Distribution

This is a distribution associated with “rare events”, events occur in given times or on defined spaces.

- traffic accidents,
- typing errors,
- customers arriving in a bank,
- telephone calls at a call center,
- document requests on a web server,
- and many other punctual phenomena where events occur independently from each other.

Definition of Poisson Distribution

Let the random variable X count the *number of events occurring in unit time interval* [0, 1], and let λ be the average number of events.

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} (x = 0, 1, 2, \cdots)$$

We say that X has a **Poisson Distribution** and write

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

The mean, variance and mgf of the Poisson are:

$$\text{E}(X) = \lambda, \ \text{D}(X) = \lambda, \ M(t) = e^{\lambda(e^t - 1)}$$

The associated R functions are dpois(x, lambda) .

Poisson Process

The **Poisson process** is a continuous-time counting process $\{X(t),\ t \geq 0\}$ that possesses the following properties:

1. The number of outcomes in non-overlapping intervals are independent.
2. The probability of two or more outcomes in a sufficiently short interval is virtually zero.
3. The probability of exactly one outcome in a sufficiently short interval or small region is proportional to the length of the interval or region.

Why we call poisson process?

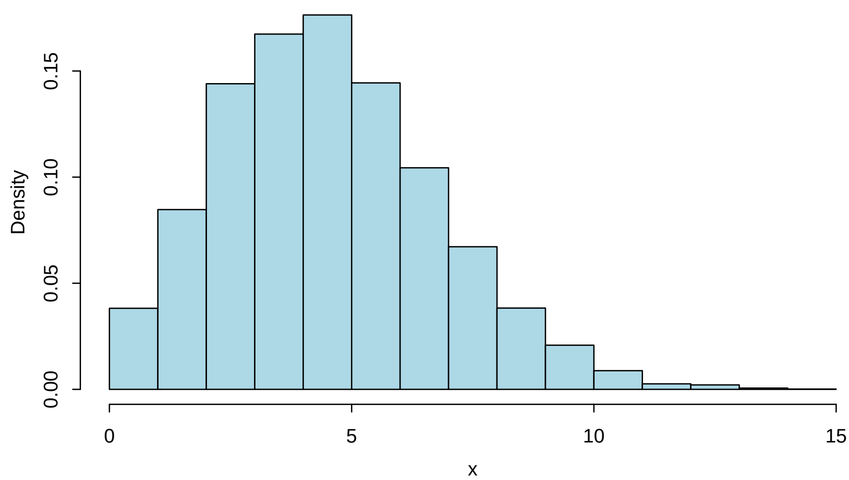
- For each $t \geq 0$, the probability distribution of $X(t)$ is a **Poisson distribution** with parameter λt. (Here $\lambda > 0$ is called the **rate** of the Poisson process),

$$P(X(t) = x|\lambda t) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}(x = 0, 1, 2, \cdots), \lambda > 0$$

- The distribution of the waiting time until the next occurrence is an **exponential distribution**.
- The occurrences are distributed uniformly on any interval of time.
- Note that $X(t)$, the total number of occurrences, has a **Poisson distribution** over the non-negative integers, whereas the location of an individual occurrence on $t \in (a, b]$ is uniform

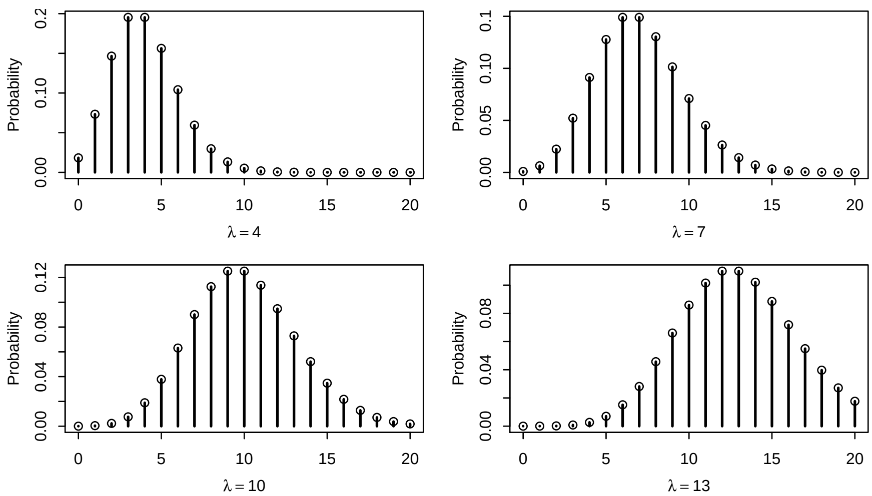
Histogam plot of Poisson with $\lambda = 5$

```
> N<-10000; lambda<-5; x<-rpois(N,lambda)
> hist(x, probability=T, col='lightblue',main="")
```



Density Plots of Poisson

```
> opar<-par(no.readonly=TRUE)
> lambda<-c(4,7,10,13);x<-0:20;par(mfrow=c(2,2))
> for(i in 1 : 4){d<-dpois(x,lambda[i])
+                 xlab <- bquote(lambda == .(lambda[i]))
+ plot(x,d,type='h',lwd=2,xlab=xlab,ylab="Probability")
+ points(x,d)};par(opar)
```



Example: World Cup Soccer

- The World Cup goals data from 1990 to 2002
- The data frame SOCCER contains three columns:
  - *CGT*: cumulative goal time in minutes (eg, the first goal was scored at the 67th minute of the first game and the second goal was scored at the 42nd minute of the second game. Consequently, the times listed in CGT for the first two goals are 67, and 132 = 90 + 42)
  - *Game*: game in which goals were scored (Totally 232 games played from 1990 to 2002)
  - *Goals*: number of goals scored in the regulation 90-minute period (Totally 575 goals)
- All of the information contained in SOCCER is indirectly available from the FIFA World Cup website:
  - http://fifaworldcup.yahoo.com/.

Example: World Cup Soccer

```
> library("PASWR2")
> head(SOCCER)

##   cgt game goals
## 1  67    1     1
## 2 132    2     2
## 3 147    3     1
## 4 258    4     2
## 5 320    5     6
## 6 355    6     3
```

```
> xtabs(~goals, data = SOCCER)

## goals
##  0  1  2  3  4  5  6  7  8
## 19 49 60 47 32 18  3  3  1
```

Are the Soccer data match a Poisson process?

Condition 1: Independent.

compute the 1 to 5 game lagged correlation coefficients

Condition 2: Obvious;

Condition 3: Problems 4.4, question 44, page 313.

```
> LAG<-sapply(1:5,function(x){
+     SOCCER$goals[x:(x + 227)]})
> round(cor(LAG), 3)

##        [,1]   [,2]   [,3]   [,4]   [,5]
## [1,]  1.000 -0.049  0.055 -0.138 -0.008
## [2,] -0.049  1.000 -0.046  0.044 -0.138
## [3,]  0.055 -0.046  1.000 -0.054  0.045
## [4,] -0.138  0.044 -0.054  1.000 -0.057
## [5,] -0.008 -0.138  0.045 -0.057  1.000
```

Are the Soccer goals data match a Poisson distribution?

```
> OBS = xtabs(~goals, data = SOCCER)
> Empir = round(OBS/sum(OBS), 3)
> TheoP=round(dpois(0:(length(OBS)-1),mean(SOCCER$goals,na.rm=TRUE)),3)
> EXP = round(TheoP * 232, 0)
> ANS = cbind(OBS, EXP, Empir, TheoP)
> ANS

##   OBS EXP Empir TheoP
## 0  19  19 0.082 0.084
## 1  49  48 0.211 0.208
## 2  60  60 0.259 0.258
## 3  47  49 0.203 0.213
## 4  32  31 0.138 0.132
## 5  18  15 0.078 0.065
## 6   3   6 0.013 0.027
## 7   3   2 0.013 0.010
## 8   1   1 0.004 0.003
```

2 Continuous Univariate Distributions

Objectives

How to choose a reasonable continuous model under a variety of physical circumstances

 Common continuous distributions

- Uniform and Normal;
- Exponential and Gamma;
- Beta, Cauchy, Lognormal and Weibull

2.1 Uniform Distribution and Normal Distribution

Uniform Distribution

The uniform distribution $X \sim \text{Unif}(a, b)$, is the probability distribution of random number selection from the continuous interval between a and b. Its density function is defined by

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

with expectation $\text{E}(X) = (a + b)/2$ and variance $\text{Var}(X) = (b - a)^2/12$

The R function is dunif(min = a, max = b).

Simulation: Inverse Transformation Method

Theorem 1. *If X is continuous random variable with cdf $F(x)$, then the transformed variable $Y = F(X) \sim \text{Unif}(0, 1)$.*

Example 2. Generate a sample of 1000 random values from a continuous distribution with $X \sim f(x) = 3x^2 (0 < x < 1)$.

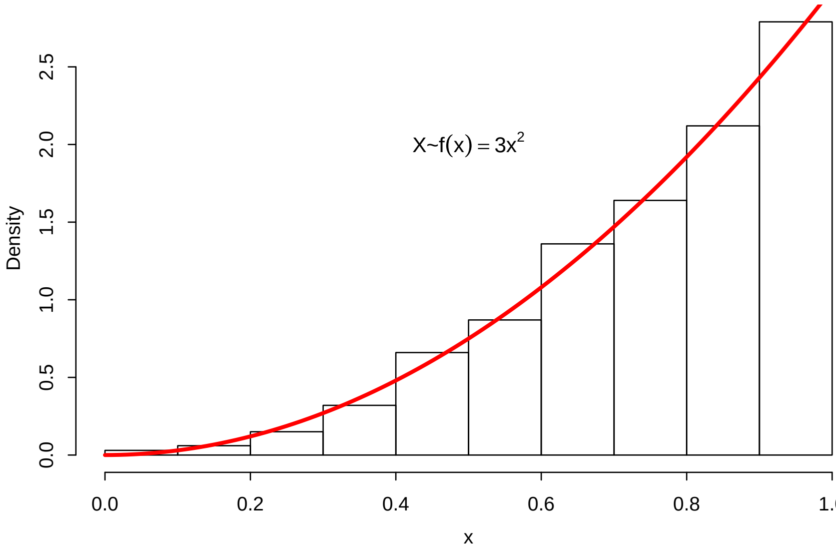
Solution: $F(x) = x^3(0 < x < 1)$, Since $U = F(X) = X^3 \sim \text{unif}(0, 1)$, $X = U^{1/3}$, hence we can first draw sample from U: $u_1, u_2, \ldots, u_n$, then

$$x_1 = u_1^{1/3},\ x_2 = u_2^{1/3}, \ldots, x_n = u_n^{1/3}$$

are samples from X.

How to generate samples from an arbitrary r.v?

```
> u<-runif(1000);x<-u^(1/3)
> hist(x,prob=TRUE,main="")
> curve(3*x^2,from=0,to=1,add=TRUE,lwd=3,col="red")
> text(0.5,2,expression({X}*"~"*{f(x)==3*x^2}))
```



Normal Distribution

A random variable X is called a normal or Gaussian distribution, denoted by $X \sim N(\mu, \sigma^2)$, if its pdf function is:

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

with $E(X) = \mu$, $Var(X) = \sigma^2$ and $M(t) = e^{\mu t + \sigma^2 t^2/2}$.

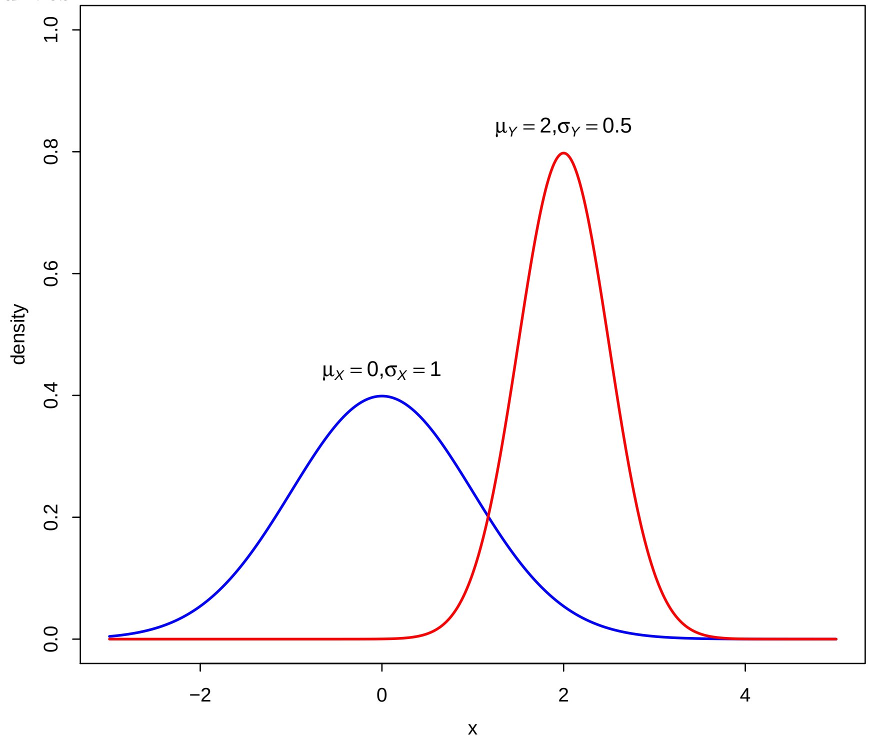
Standard normal $Z \sim N(0,1)$: If $\mu = 0, \sigma = 1$

- The cdf of $N(0,1)$ is denoted by $\Phi(x)$
- Notation: z_α denotes the α lower quartile of the standard normal distribution, i. e. $P(Z < z_\alpha) = \alpha$

Useful R functions for Normal

- `dnorm(x,mean=0,sd=1)`
  - Density function.
- `pnorm(x,mean=0,sd=1,lower.tail=TRUE)`
  - Probabilty $P(X \leq x)$.
- `qnorm(p,mean=0,sd=1,lower.tail=TRUE)`
  - Lower quartile x_p such that $P(X \leq x_p) = p$
- `rnorm(n,mean=0,sd=1)`
  - Generate a sample of size n

Normal Curves



Some Useful Facts:

- If $X \sim N(\mu, \sigma^2)$, then

$$
\begin{aligned}
P(\mu - 1\sigma < X < \mu + 1\sigma) &= 68\% \\
P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) &= 95\% \\
P(\mu - 3\sigma < X < \mu + 3\sigma) &= 99.73\%
\end{aligned}
$$

- In words: the chance that X will be within 1σ (i.e. one standard deviation) of its mean is about 68%, and the chance that it will be within 2σ of its mean is about 95%

- Standardization: If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Example: The 1987 crash

How extreme was the 1987 crash of − 21.76%?

1. Prior to the October, 1987 crash SP500 monthly returns were 1.2% with a risk/volatility of 4.3%.

$$X \sim N(0.012, 0.043^2)$$

2. Standardize:

$$Z = \frac{X - 0.012}{0.043} = \frac{-0.2176 - 0.012}{0.043} = -5.27$$

That is a 5-sigma event!

(The probability outside 3-sigma is almost zero)

2.2 Exponential Distribution and Gamma Distribution

Exponential Distribution

The exponential distribution describes the arrival time of a randomly recurring independent event sequence.

If λ is the rate of event recurrence, then $X \sim \text{Exp}(\lambda)$ with probability density function:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

which with mean $\text{E}(X) = 1/\lambda$, variance $1/\lambda^2$ and $M(t) = (1 - \lambda^{-1}t)^{-1}$ for $t < \lambda$.

R functions: dexp(x, rate = 1),...

Memoryless

Assume lifetime $X \sim \text{Exp}(\lambda)$, it can be shown that

$$P(X > s + t | X > s) = P(X > t)$$

Exponential distribution and Poisson Process

The exponential distribution is closely related to the Poisson distribution.

- If customers arrive at a store according to a Poisson process with rate λ and
- if Y counts the number of customers that arrive in the time interval$[0, t)$, then $Y \sim \text{Pois}(\lambda t)$

Let T be the length of times between two adjacent customers arrive. It can be shown that $T \sim \text{Exp}(\lambda)$.

Proof:

$$P(T > t) = P(\text{no event in } [0, t)) = P(Y = 0) = e^{-\lambda t}$$

implies the cdf of T is $F(t) = 1 - e^{-\lambda t}$ $(t > 0)$, hence $T \sim \text{Exp}(\lambda)$.

Soccer: Are the Inter-goal times exponential?

Mean and standard deviation for the time between goals

Let $T_i = cgt_{i+1} - cgt_i$ $(i = 1, 2, \ldots, 574)$, $\lambda = 575/232$ goals/game, the average time (in minutes) before a goal is scored is $90/\lambda$ =36.313 minutes.

```
> inter.times<-with(data=SOCCER,cgt[2:575]-cgt[1:574])
> MEAN <- mean(inter.times)
> SD <- sd(inter.times)
> c(MEAN = MEAN, SD = SD)

##     MEAN       SD
## 36.24042 36.67138
```

Soccer: Are the Inter-goal times exponential?

Is it reasonable to model the time between goals with the exponential distribution?

Split the data into discrete categories with bins =10, then compare the observed and theoretical values.

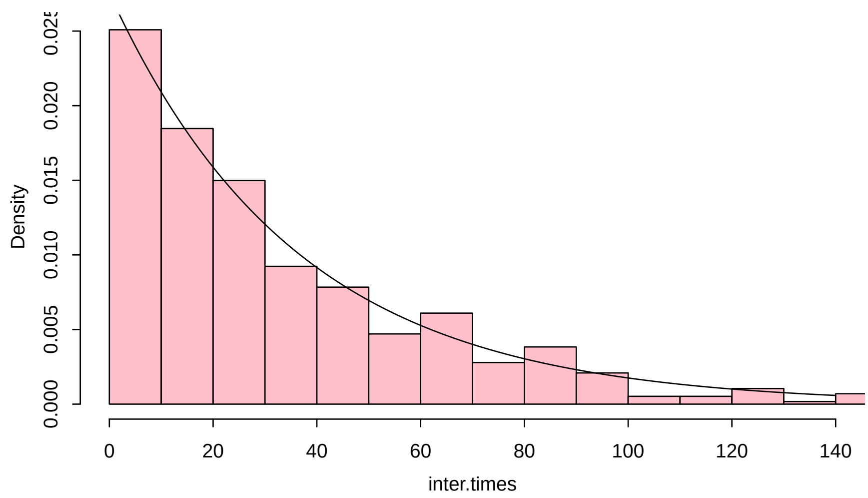
```
rate <- (575/232)*(1/90) # rate = lambda*t
nit <- sum(!is.na(inter.times)) # number of inter.times
OBS<-xtabs(~cut(inter.times,breaks=c(seq(0,130,10),310)))
EmpiP<-round(OBS/nit, 3)
TheoP<-round(c((pexp(seq(10, 130, 10),rate) -
         pexp(seq(0,120,10),rate)),(1-pexp(130,rate))),3)
EXP <-round(TheoP*nit, 0)
ANS <-cbind(OBS, EXP, EmpiP, TheoP)
ANS
```

Soccer: Are the Inter-goal times exponential?

```
##             OBS EXP EmpiP TheoP
## (0,10]      144 138 0.251 0.241
## (10,20]     106 105 0.185 0.183
## (20,30]      86  80 0.150 0.139
## (30,40]      53  60 0.092 0.105
## (40,50]      45  46 0.078 0.080
## (50,60]      27  35 0.047 0.061
## (60,70]      35  26 0.061 0.046
## (70,80]      16  20 0.028 0.035
## (80,90]      22  15 0.038 0.027
## (90,100]     12  11 0.021 0.020
## (100,110]     3   9 0.005 0.015
## (110,120]     3   7 0.005 0.012
## (120,130]     6   5 0.010 0.009
## (130,310]    16  16 0.028 0.028
```

Soccer: Are the Inter-goal times exponential?

```
> hist(inter.times,breaks=seq(0,310,10),col="pink",xlim=c(0,140),
+      prob=TRUE,main="")
> xt<-seq(0,140,0.01);ft<-dexp(xt,rate);lines(xt,ft,type="l")
```



Check the lack of memory property

Empirically, with mean 36.313,

$$P(T > 10) = 1 - P(T \le 10) = 1 - 144/574 = 430/574 = 0.749$$

$$P(T > 20|T > 10) = (574 - 144 - 106)/(574 - 144) = 0.754$$

They are rather close.

Gamma Function

Gamma function:

$$\Gamma(\alpha) = \int_0^{+\infty} x^\alpha e^{-x} \mathrm{d}x,\ \alpha > 0$$

Properties:

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
2. For any positive integer n, $\Gamma(n) = (n - 1)!$
3. $\Gamma(1/2) = \sqrt{\pi}$

Gamma Distribution

The gamma distribution is often used to model the waiting time until the α^{th} event in a Poisson process.

The Gamma distribution $X \sim \text{gamma}(\alpha, \lambda)$ or $X \sim \Gamma(\alpha, \lambda)$ with parameters `shape` $= \alpha$ and rate $= \lambda$ (scale $= 1/\lambda$) has density

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} (x > 0)$$

with $\mathrm{E}(X) = \alpha/\lambda$, $\mathrm{Var}(X) = \alpha/\lambda^2$, and

$$M(t) = (1 - \lambda^{-1}t)^{-\alpha} \text{ for } t < \lambda$$

R functions: dgamma(x, shape, rate = 1),...

Motivation

- if X measures the length of time until the first event occurs in a Poisson process with rate λ, then $X \sim \text{Exp}(\text{rate} = \lambda)$.
- If we let Y measure the length of time until the αth event occurs, then $Y \sim \Gamma(\text{shape} = \alpha,\ \text{rate} = \lambda)$.
- hence $\Gamma(1, \lambda) = \text{Exp}(\lambda)$.
- When α is an integer this distribution is also known as the **Erlang distribution**.
- Given $X \sim \Gamma(\alpha, \lambda)$. The sampling distribution of $\bar{X}$ for a random sample of size n is $\Gamma(n\alpha, n\lambda)$.

2.3 Beta, Cauchy, Lognormal and Weibull

Beta function

Beta function:

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx \ (a > 0, b > 0)$$

Relationship between Beta function and Gamma function:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Properties:

1. $B(a,b) = B(b,a)$
2. $B(a,b) = B(a,b+1) + B(a+1,b)$
3. $B(a+1,b) = B(a,b)\frac{a}{a+b}$
4. $B(a,b+1) = B(a,b)\frac{b}{a+b}$

Beta Distribution

The Beta distribution $X \sim \text{Beta}(\alpha,\beta)$ with parameters shape1 $= \alpha$ and shape2 $= \beta$ has density

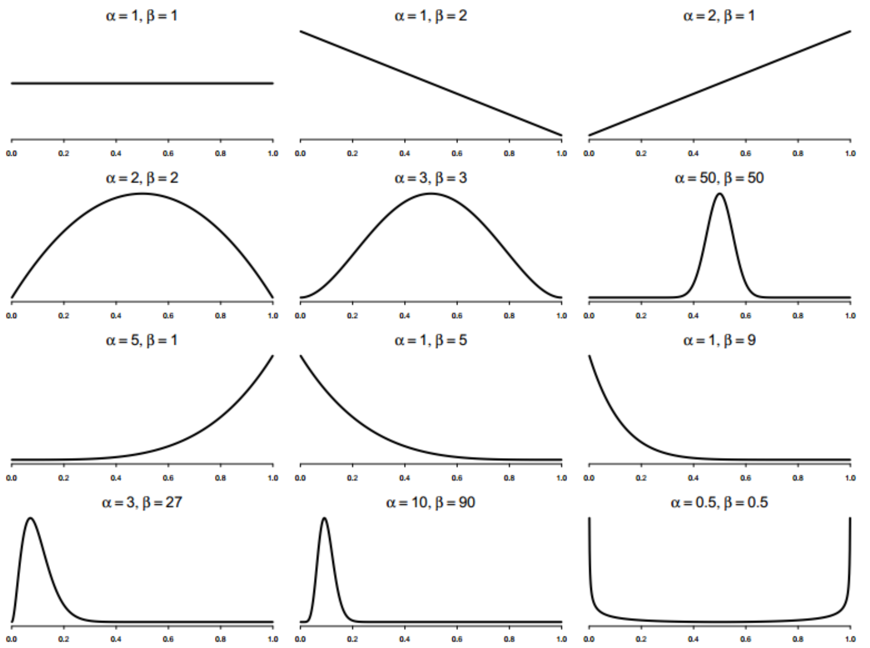
$$f(x) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}(0 < x < 1)$$

with

$$\text{E}(X) = \frac{\alpha}{\alpha+\beta}, \ \text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- This is a generalization of the continuous uniform distribution: $\text{Beta}(\alpha = 1, \beta = 1) = \text{Unif}(0,1)$.
- This distribution comes up a lot in Bayesian statistics because it is a good model for one's prior beliefs about a population **proportion** p, $0 \leq p \leq 1$.

Beta distribution under different parameters



R functions

The associated R functions are:

- `dbeta(x, shape1, shape2, log = FALSE)`
- `pbeta(q, shape1, shape2, lower.tail = TRUE)`
- `qbeta(p, shape1, shape2, lower.tail = TRUE)`
- `rbeta(n, shape1, shape2)`

Cauchy Distribution

This is a special case of the Student's t-distribution.
The **Cauchy Distribution** has the following PDF

$$f(x) = \frac{1}{\beta\pi}\left[1 + \left(\frac{x-m}{\beta}\right)^2\right]^{-1}, \quad -\infty < x < \infty$$

We write $X \sim \text{cauchy}(\text{location} = m, \text{scale} = \beta)$.
The associated R function is

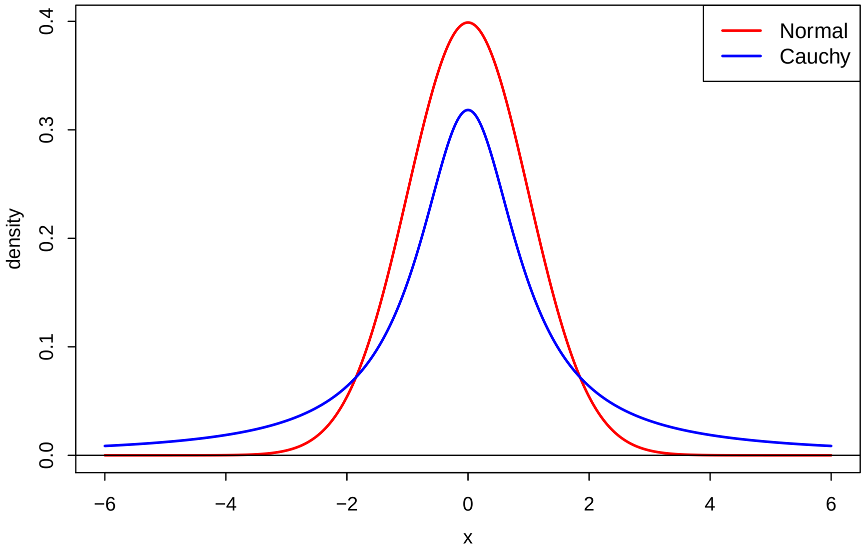
$$\text{dcauchy(x, location} = 0\text{, scale} = 1)$$

Properties

1. It is easy to see that a `cauchy(location = 0, scale = 1)` distribution is the same as a `t(df = 1)` distribution, that is $\text{cauchy}(0, 1) = t(1)$.
2. The `cauchy` distribution looks like a `normal` distribution but with very **heavy tails**.

3. The mean (and variance) **do not exist**, that is, they are infinite.
4. The median is represented by the `location` parameter, and the `scale` parameter influences the spread of the distribution about its median.

```
> x<-seq(-6,6,length=501)
> plot(x,dnorm(x),ylab='density',type='l',lwd=2,col='red')
> lines(x,dcauchy(x),lwd=2,col='blue'); abline(h=0)
> legend("topright",legend=c('Normal','Cauchy'),
+        col=c('red','blue'),lty=1,lwd=2)
```



Lognormal Distribution

Let $Y \sim N(\mu, \sigma^2)$ and $X = e^Y$. Then the density of X is **lognormal**.
We write $X \sim \text{lnorm}(\mu, \sigma^2)$ with pdf

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right], \quad x > 0$$

The associated R function is:
`dlnorm(x, meanlog = 0, sdlog = 1),...`

Weibull Distribution

We write $X \sim \text{Weibull}(\alpha, \beta)$ with pdf

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(\frac{x}{\beta}\right)^{\alpha}, \quad x > 0$$

The associated R function is:
`dweibull(x, shape, scale = 1),...`