Probability and Statistics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

CHAPTER 3: PARAMETRIC FAMILIES OF UNIVARIATE DISTRIBUTIONS

1 Why do we need distributions?

1.1 Some practical uses of probability distributions

1.2 Related distributions

1.3 Families of probability distributions

2 Discrete distributions

2.1 Introduction

2.2 Discrete uniform distributions

2.3 Bernoulli and binomial distribution

2.4 Hypergeometric distribution

2.5 Poisson distribution

3 Continuous distributions

3.1 Introduction

3.2 Uniform or rectangular distribution

3.3 Normal distribution

3.4 Exponential and gamma distribution

3.5 Beta distribution

4 Where discrete and continuous distributions meet

4.1 Approximations

4.2 Poisson and exponential relationships

4.3 Deviations from the ideal world ?

4.3.1 Mixtures of distributions

4.3.2 Truncated distributions

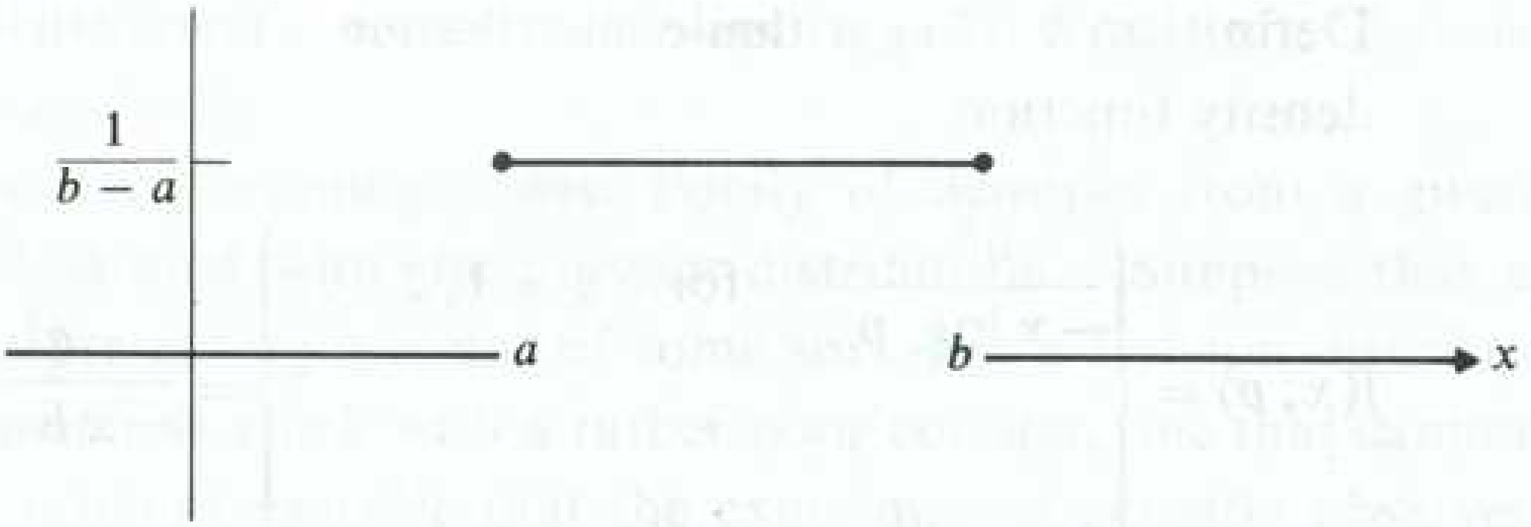
3 Continuous distributions

3.1 Introduction

Distribution	Probability Density Function $f(x)$	Mean	Variance
Uniform $U(\alpha, \beta)$	$\frac{1}{(\beta - \alpha)}, \quad \alpha \leq x \leq \beta$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$
Normal $N(\mu, \sigma)$	$\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty$	μ	σ^2
Exponential $\text{Exp}(\beta)$	$\frac{1}{\beta} e^{\ \frac{1}{\beta} x}, \quad 0 < x < \infty$	β	β^2
Lognormal $\text{lognormal}(\alpha, \beta)$	$\frac{1}{\sqrt{2\pi}\beta} x^{-1} e^{-(\ln x - \alpha)^2/2\beta^2}, \quad 0 < x < \infty$	$e^{\alpha + \beta^2/2}$	$e^{2\alpha + \beta^2}\left[e^{\beta^2} - 1\right]$

Distribution	Probability Density Function $f(x)$	Mean	Variance
Gamma $\text{Gamma}(\alpha, \beta)$	$\frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty$	$\alpha\beta$	$\alpha\beta^2$
Beta $\text{Beta}(\alpha, \beta)$	$\frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 \quad x)^{\beta-1}, \quad 0 \le x \le 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha-\beta+1)}$
Weibull $\text{Weibull}(\alpha, \beta)$	$(\alpha\beta)x^{\beta} e^{-\alpha x^{\beta}}, \quad 0 \le \infty$	$\alpha^{-1/\beta}\Gamma(1+\frac{1}{\beta})$	$\alpha^{-2/\beta}\left\{\Gamma(1+\frac{2}{\beta}) - \left[\Gamma(1+\frac{1}{\beta})\right]^2\right\}$

3.2 Uniform or rectangular distribution



Definition **Uniform distribution** If the probability density function of a random variable X is given by

$$f_X(x) = f_X(x; a, b) = \frac{1}{b-a} I_{[a,b]}(x), \qquad 1)$$

Theorem If X is uniformly distributed over $[a, b]$, then

$$\mathscr{E}[X] = \frac{a+b}{2}, \quad \text{var}[X] = \frac{(b-a)^2}{12}, \quad \text{and} \quad m_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

Proof

$$\mathscr{E}[X] = \int_a^b x \frac{1}{b-a}\, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

$$\text{var}[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = \int_a^b x^2 \frac{1}{b-a}\, dx - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

$$m_X(t) = \mathscr{E}[e^{tX}] = \int_a^b e^{tx} \frac{1}{b-a}\, dx = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

////

Common statistics

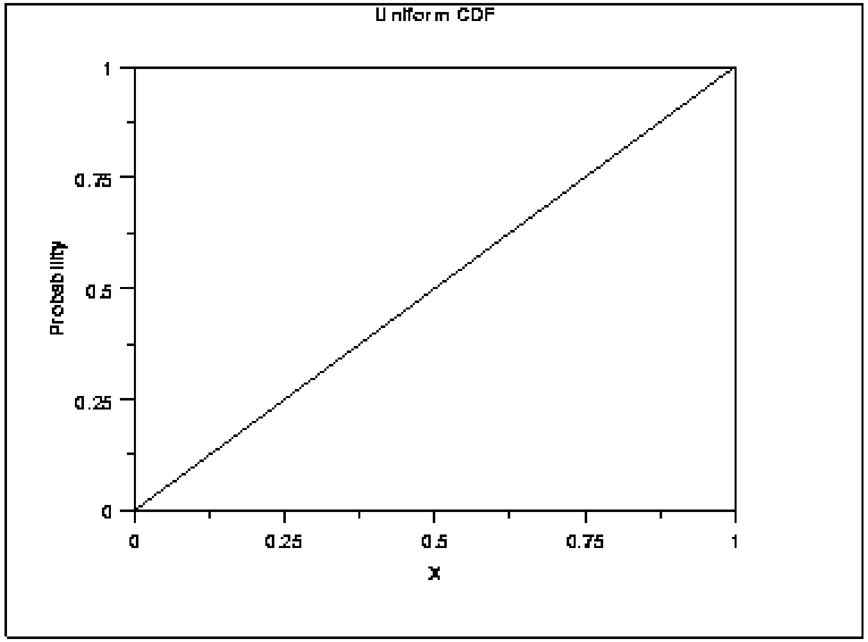
Mean	(A + B)/2
Median	(A + B)/2
Range	B - A
Standard Deviation	$\sqrt{\frac{(B-A)^2}{12}}$
Coefficient of Variation	$\frac{(B-A)}{\sqrt{3}(B+A)}$
Skewness	0
Kurtosis	9/5

Cumulative distribution function

- The formula for the cumulative distribution function of the uniform distribution is

$$F(x) = x \qquad \text{for } 0 \leq x \leq 1$$

- The following is the plot of the uniform cumulative distribution function.

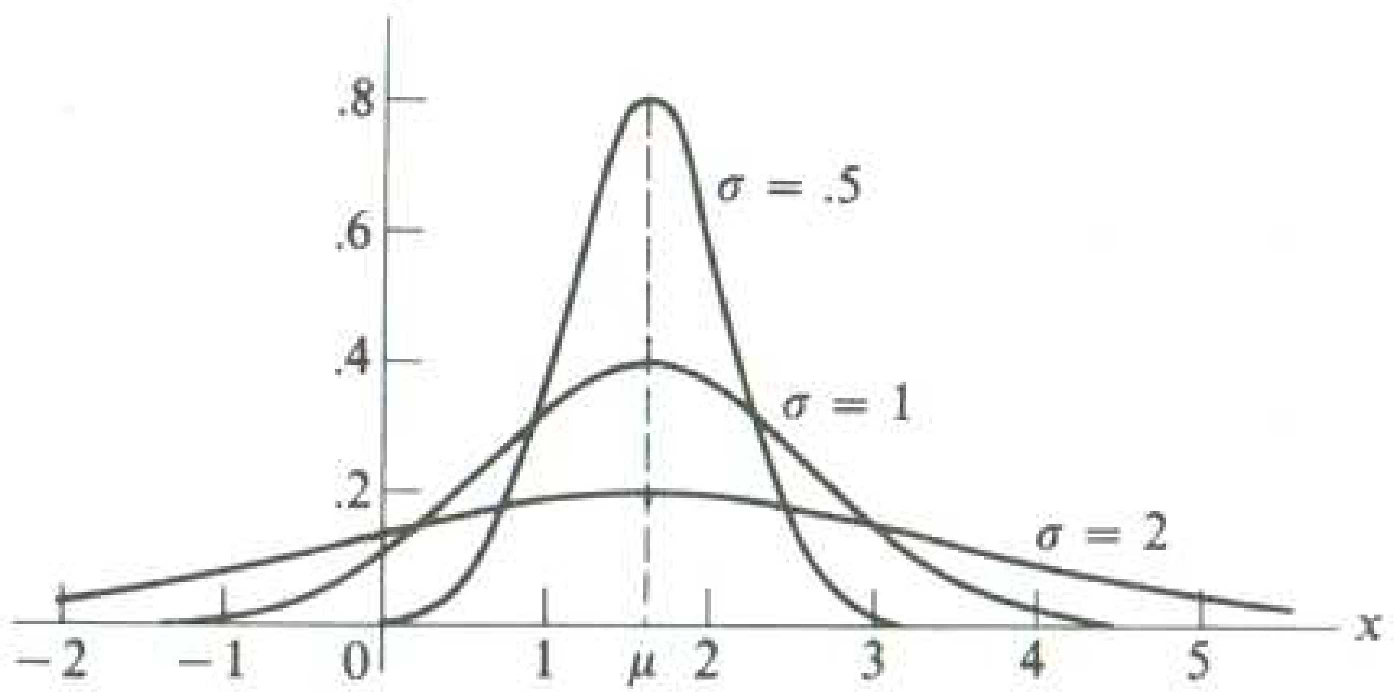


Example

- It provides a useful model for a few random phenomena. For instance, if it is known that the values of some random variable *X* can only be in a finite interval, say *[a,b]*, and if one assumes that any two subintervals of *[a,b]* of equal length have the same probability of containing *X*, then *X* has a uniform distribution over the interval [a,b].
- When it is mentioned that a **random number** is taken from the interval [0,1], it means that a value is obtained from a uniformly distributed random variable over the interval [0,1].

EXAMPLE If a wheel is spun and then allowed to come to rest, the point on the circumference of the wheel that is located opposite a certain fixed marker could be considered the value of a random variable X that is uniformly distributed over the circumference of the wheel. One could then compute the probability that X will fall in any given arc. ////

3.3 Normal distribution



- The mode occurs at μ
- The inflection points occur at μ±σ

Definition **Normal distribution** A random variable X is defined to be *normally* distributed if its density is given by

$$f_X(x) = f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},$$

where the parameters μ and σ satisfy $-\infty < \mu < \infty$ and $\sigma > 0$. Any distribution defined by a density function given in Eq. is called a *normal distribution*. ////

Remark:

We have used the symbols μ and σ^2 to represent the parameters because these parameters turn out, as we shall see, to be the mean and variance, respectively, of the distribution.

Theorem If X is a normal random variable,

$$\mathscr{E}[X] = \mu, \qquad \text{var}\,[X] = \sigma^2, \qquad \text{and} \qquad m_X(t) = e^{\mu t + \sigma^2 t^2/2}.$$

Proof

$$\begin{aligned} m_X(t) = \mathscr{E}[e^{tX}] &= e^{t\mu}\mathscr{E}[e^{t(X-\mu)}] \\ &= e^{t\mu}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{t(x-\mu)} e^{-(1/2\sigma^2)(x-\mu)^2}\, dx \\ &= e^{t\mu}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1/2\sigma^2)[(x-\mu)^2 - 2\sigma^2 t(x-\mu)]}\, dx. \end{aligned}$$

If we complete the square inside the bracket, it becomes

$$\begin{aligned} (x-\mu)^2 - 2\sigma^2 t(x-\mu) &= (x-\mu)^2 - 2\sigma^2 t(x-\mu) + \underline{\sigma^4 t^2 - \sigma^4 t^2} \\ &= (x-\mu-\sigma^2 t)^2 - \sigma^4 t^2, \end{aligned}$$

and we have

$$m_X(t) = e^{t\mu} e^{\sigma^2 t^2/2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu-\sigma^2 t)^2/2\sigma^2}\, dx.$$

The integral together with the factor $1/\sqrt{2\pi}\sigma$ is necessarily 1 since it is the area under a normal distribution with mean $\mu + \sigma^2 t$ and variance σ^2. Hence,

$$m_X(t) = e^{\mu t + \sigma^2 t^2/2}.$$

On differentiating $m_X(t)$ twice and substituting $t = 0$, we find

$$\mathscr{E}[X] = m_X'(0) = \mu$$

and

$$\text{var}\,[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = m_X''(0) - \mu^2 = \sigma^2,$$

thus justifying our use of the symbols μ and σ^2 for the parameters. ////

Common statistics

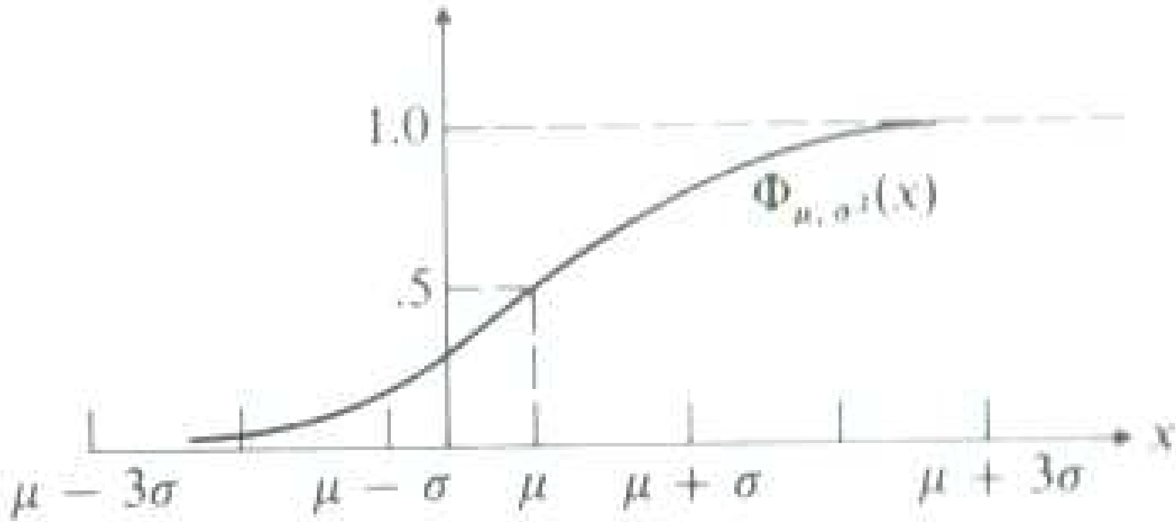
Mean	The location parameter μ.
Median	The location parameter μ.
Mode	The location parameter μ.
Range	Infinity in both directions.
Standard Deviation	The scale parameter σ.
Coefficient of Variation	σ/μ
Skewness	0
Kurtosis	3

Cumulative distribution function

If random variable X is normally distributed with mean μ and variance σ^2, we will write $X \sim N(\mu, \sigma^2)$. We will also use the notation $\phi_{\mu,\sigma^2}(x)$ for the density of $X \sim N(\mu, \sigma^2)$ and $\Phi_{\mu,\sigma^2}(x)$ for the cumulative distribution function.

If the normal random variable has mean 0 and variance 1, it is called a *standard* or *normalized* normal random variable. For a standard normal random variable the subscripts of the density and distribution function notations are dropped; that is,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \text{and} \qquad \Phi(x) = \int_{-\infty}^{x} \phi(u)\, du.$$


1.0
.5
$\Phi_{\mu,\sigma^2}(x)$
x
$\mu - 3\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 3\sigma$

- The following theorem shows that we can find the probability that a normally distributed random variable, with mean μ and variance σ^2, falls in any interval in terms of the standard normal cumulative distribution function (for which tables exist)

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

$$P[a < X < b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

PROOF

$$\begin{aligned} P[a < X < b] &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}\,dx \\ &= \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,dz \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

////

Remark

$$\Phi(x) = 1 - \Phi(-x).$$

Example

- The normal distribution appears to be a reasonable model of the behavior of certain random phenomena.
- It is also the limiting distribution (limiting form) of many other probability distributions. Limiting for what?

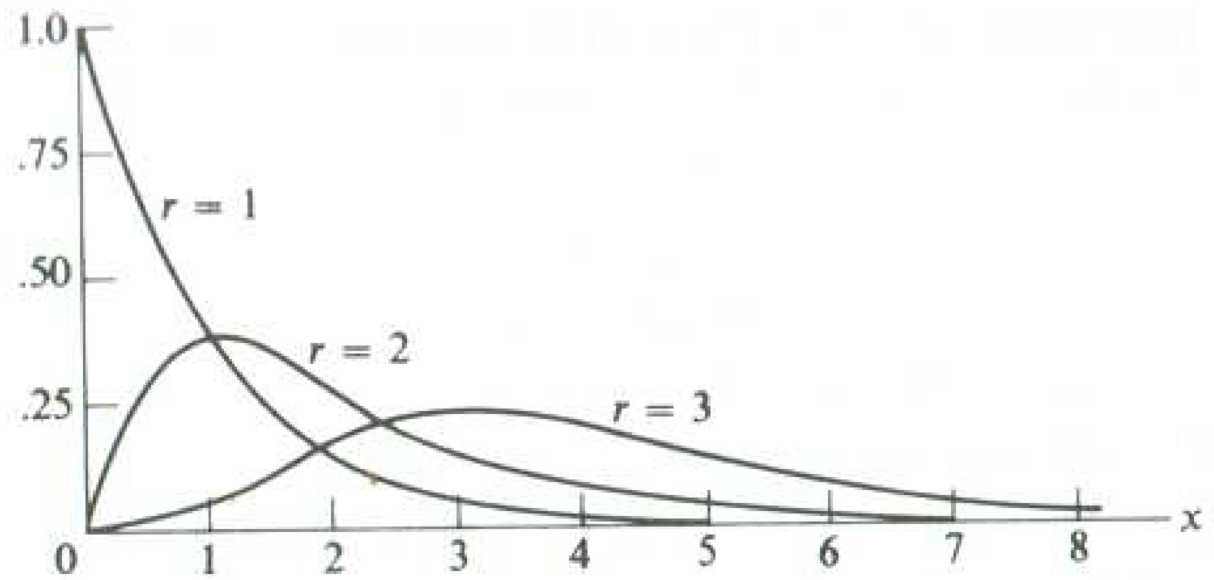
EXAMPLE Suppose that an instructor assumes that a student's final score is the value of a normally distributed random variable. If the instructor decides to award a grade of A to those students whose score exceeds $\mu + \sigma$, a B to those students whose score falls between μ and $\mu + \sigma$, a C if a score falls between $\mu - \sigma$ and μ, a D if a score falls between $\mu - 2\sigma$ and $\mu - \sigma$, and an F if the score falls below $\mu - 2\sigma$, then the proportions of each grade given can be calculated. For example, since

$$P[X > \mu + \sigma] = 1 - P[X < \mu + \sigma] = 1 - \Phi\left(\frac{\mu + \sigma - \mu}{\sigma}\right)$$

$$= 1 - \Phi(1) \approx .1587,$$

one would expect 15.87 percent of the students to receive A's. ////

3.4 Exponential and gamma distribution



(gamma densities; λ=1)

Definition **Exponential distribution** If a random variable X has a density given by

$$f_X(x;\lambda) = \lambda e^{-\lambda x} I_{[0,\infty)}(x),$$

where $\lambda > 0$, then X is defined to have an (negative) *exponential distribution*. ////

Definition **Gamma distribution** If a random variable X has density given by

$$f_X(x;r,\lambda) = \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} I_{[0,\infty)}(x),$$

where $r > 0$ and $\lambda > 0$, then X is defined to have a *gamma distribution*. $\Gamma(\cdot)$ is the gamma function ////

The gamma function

- The gamma function denoted by $\Gamma(\dot{})$ is defined by

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx$$

for t>O.

- Integration by parts yields

$$\Gamma(t+1) = t\Gamma(t)$$

- With t=n (an integer)

$$\Gamma(n+1) = n!$$

- Note: $\Gamma(1/2) = \sqrt(\pi)$

Theorem If X has an exponential distribution, then

$$\mathscr{E}[X] = \frac{1}{\lambda}, \qquad \operatorname{var}[X] = \frac{1}{\lambda^2}, \qquad \text{and} \qquad m_X(t) = \frac{\lambda}{\lambda - t} \qquad \text{for} \qquad t < \lambda.$$

Theorem If X has a gamma distribution with parameters r and λ, then

$$\mathscr{E}[X] = \frac{r}{\lambda}, \qquad \operatorname{var}[X] = \frac{r}{\lambda^2}, \qquad \text{and} \qquad m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r \qquad \text{for } t < \lambda.$$

Proof

$$m_X(t) = \mathscr{E}[e^{tX}]$$
$$= \int_0^\infty \frac{\lambda^r}{\Gamma(r)} e^{tx} x^{r-1} e^{-\lambda x}\, dx$$
$$= \left(\frac{\lambda}{\lambda - t}\right)^r \int_0^\infty \frac{(\lambda - t)^r}{\Gamma(r)} x^{r-1} e^{-(\lambda - t)x}\, dx = \left(\frac{\lambda}{\lambda - t}\right)^r.$$
$$m'_X(t) = r\lambda^r(\lambda - t)^{-r-1}$$

and

$$m''_X(t) = r(r+1)\lambda^r(\lambda - t)^{-r-2};$$

hence

$$\mathscr{E}[X] = m'_X(0) = \frac{r}{\lambda}$$

and

$$\text{var}\,[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2$$
$$= m''_X(0) - \left(\frac{r}{\lambda}\right)^2 = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \frac{r}{\lambda^2}. \qquad ////$$

Common statistics for the exponential distribution with location parameter zero

Mean	β
Median	$\beta \ln 2$
Mode	Zero
Range	Zero to plus infinity
Standard Deviation	β
Coefficient of Variation	1
Skewness	2
Kurtosis	9

where $\beta=1/\lambda$.

Common statistics for the standard gamma distribution

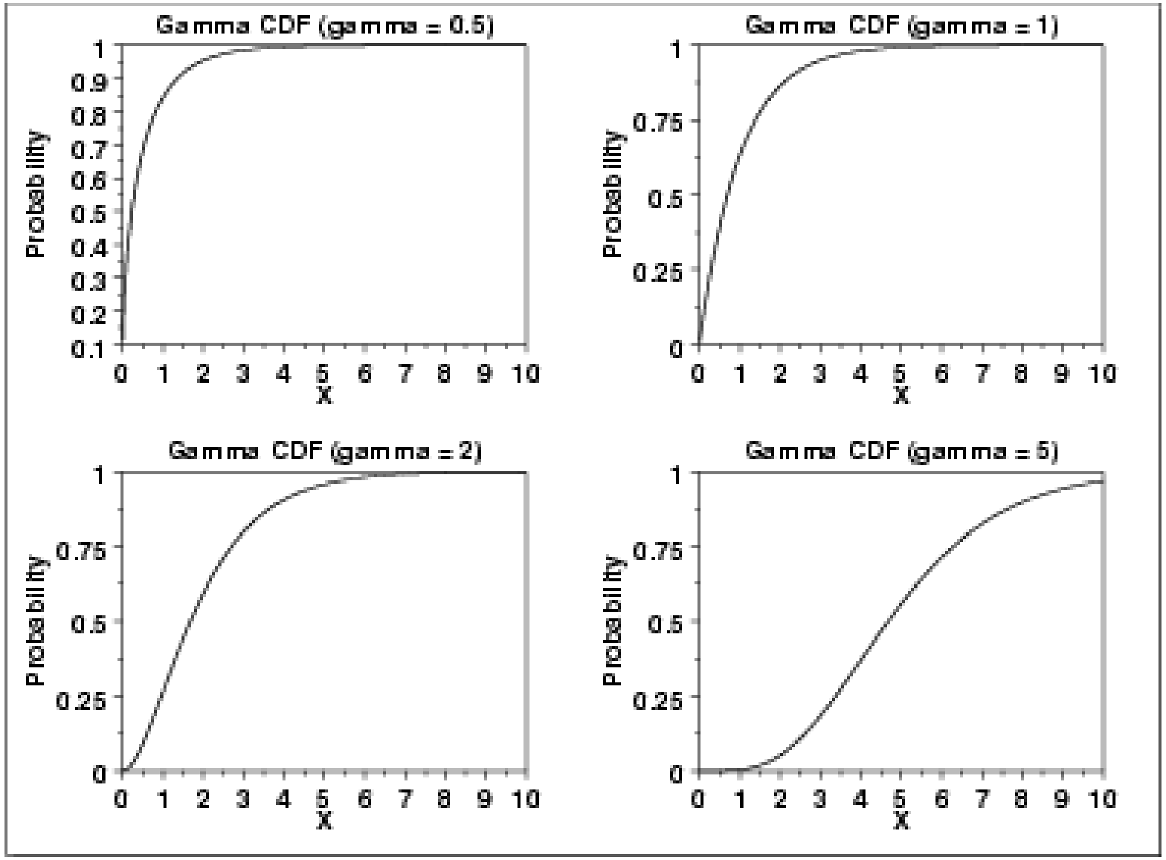
- The formula for the cumulative distribution function of the gamma distribution is

$$F(x) = \frac{\Gamma_x(\gamma)}{\Gamma(\gamma)} \qquad x \geq 0; \gamma > 0$$

where Γ is the gamma function defined above and $\Gamma_x(a)$ is the incomplete gamma function. The incomplete gamma function has the formula

$$\Gamma_x(a) = \int_0^x t^{a-1} e^{-t} dt$$

- The following is the plot of the gamma cumulative distribution function with the shape parameter gamma= r.

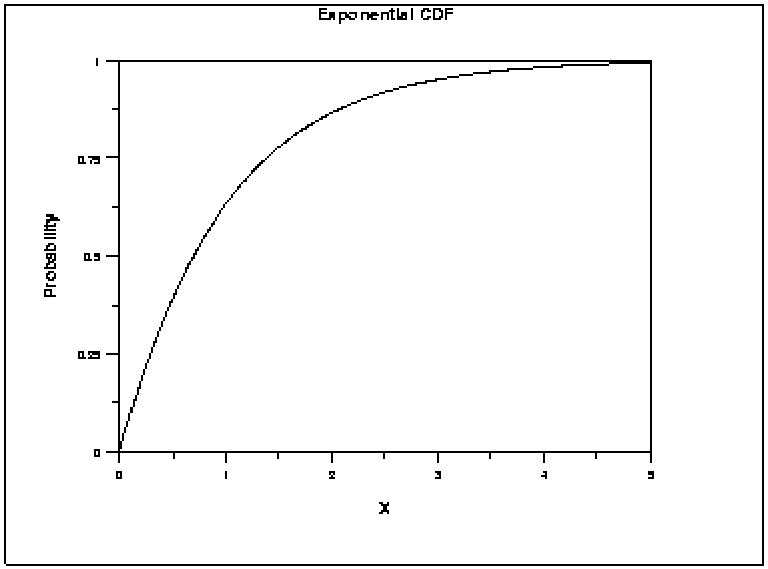


Cumulative distribution function

- The formula for the cumulative distribution function of the exponential distribution is

$$F(x) = 1 - e^{-x/\beta} \qquad x \geq 0; \beta > 0$$

- The following is the plot of the exponential cumulative distribution function.



Example

- The exponential distribution has been used as a model for lifetimes of various things.
- When we introduced the Poisson distribution, we talked about the occurrence of certain events in time or space
- The length of the time interval between successive happenings can be shown to have an exponential distribution provided that the number of happenings in a fixed time interval is Poisson.
- Also, if we assume that the number of events in a fixed time interval is Poisson distributed, then the length of time between time 0 and the instant when the r-th happening occurs, can be shown to follow a gamma distribution

Remarks

- The exponential is a special case of the gamma *(r=1)*
- The sum of independent identically distributed exponential random variables is gamma-distributed (see later)
- If a random variable X has an exponential distribution, with parameter λ, then

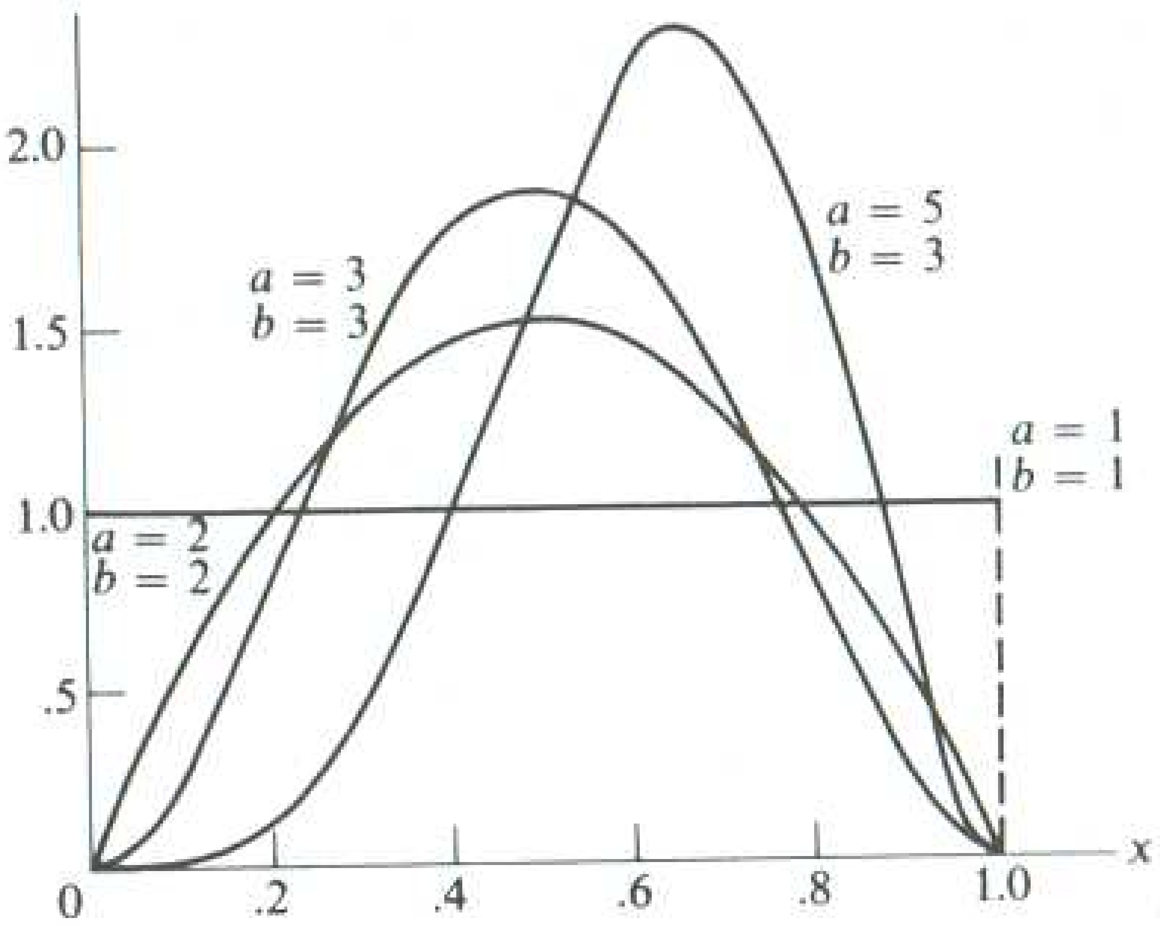
$$P[X > a + b \mid X > a] = P[X > b], \qquad \text{for } a > 0 \text{ and } b > 0.$$

PROOF $$P[X > a + b \mid X > a] = \frac{P[X > a + b]}{P[X > a]} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}}$$

$$= e^{-\lambda b} = P[X > b]. \qquad ////$$

- Let X denote the lifetime of a given component, then in words, the above results states that the conditional probability that the component will last a+b time units, given that it has lasted a time units already, is the same as its initial probability of lasting b time units.
- Hence, an “old” functioning component has the same lifetime distribution as a “new” functioning component OR the component is not subject to fatigue or wear.

3.5 Beta distribution



Definition **Beta distribution** If a random variable X has a density given by

$$f_X(x) = f_X(x; a, b) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1} I_{(0,1)}(x),$$

where $a > 0$ and $b > 0$, then X is defined to have a *beta distribution*. ////

Theorem If X is a beta-distributed random variable, then

$$\mathscr{E}[X] = \frac{a}{a+b} \qquad \text{and} \qquad \text{var}[X] = \frac{ab}{(a+b+1)(a+b)^2}.$$

Proof

$$\begin{aligned}\mathscr{E}[X^k] &= \frac{1}{B(a,b)}\int_0^1 x^{k+a-1}(1-x)^{b-1}\,dx \\ &= \frac{B(k+a,b)}{B(a,b)} = \frac{\Gamma(k+a)\Gamma(b)}{\Gamma(k+a+b)}\cdot\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\ &= \frac{\Gamma(k+a)\Gamma(a+b)}{\Gamma(a)\Gamma(k+a+b)};\end{aligned}$$

hence,

$$\mathscr{E}[X] = \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+1)} = \frac{a}{a+b},$$

and

$$\begin{aligned}\text{var}\,[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 &= \frac{\Gamma(a+2)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+2)} - \left(\frac{a}{a+b}\right)^2 \\ &= \frac{(a+1)a}{(a+b+1)(a+b)} - \left(\frac{a}{a+b}\right)^2 = \frac{ab}{(a+b+1)(a+b)^2}. \qquad ////\end{aligned}$$

Common statistics

Mean	$\frac{a}{a+b}$
Median	$I_{0.5}^{-1}(a,b)$ no closed form
Mode	$\frac{a-1}{a+b-2}$ for $a > 1, b > 1$
Standard deviation	$\frac{ab}{(a+b)^2(a+b+1)}$
Skewness	$\frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$
Kurtosis	

Cumulative distribution function

The cumulative distribution function is

$$F(x;\alpha,\beta) = \frac{\mathrm{B}_x(\alpha,\beta)}{\mathrm{B}(\alpha,\beta)} = I_x(\alpha,\beta)$$

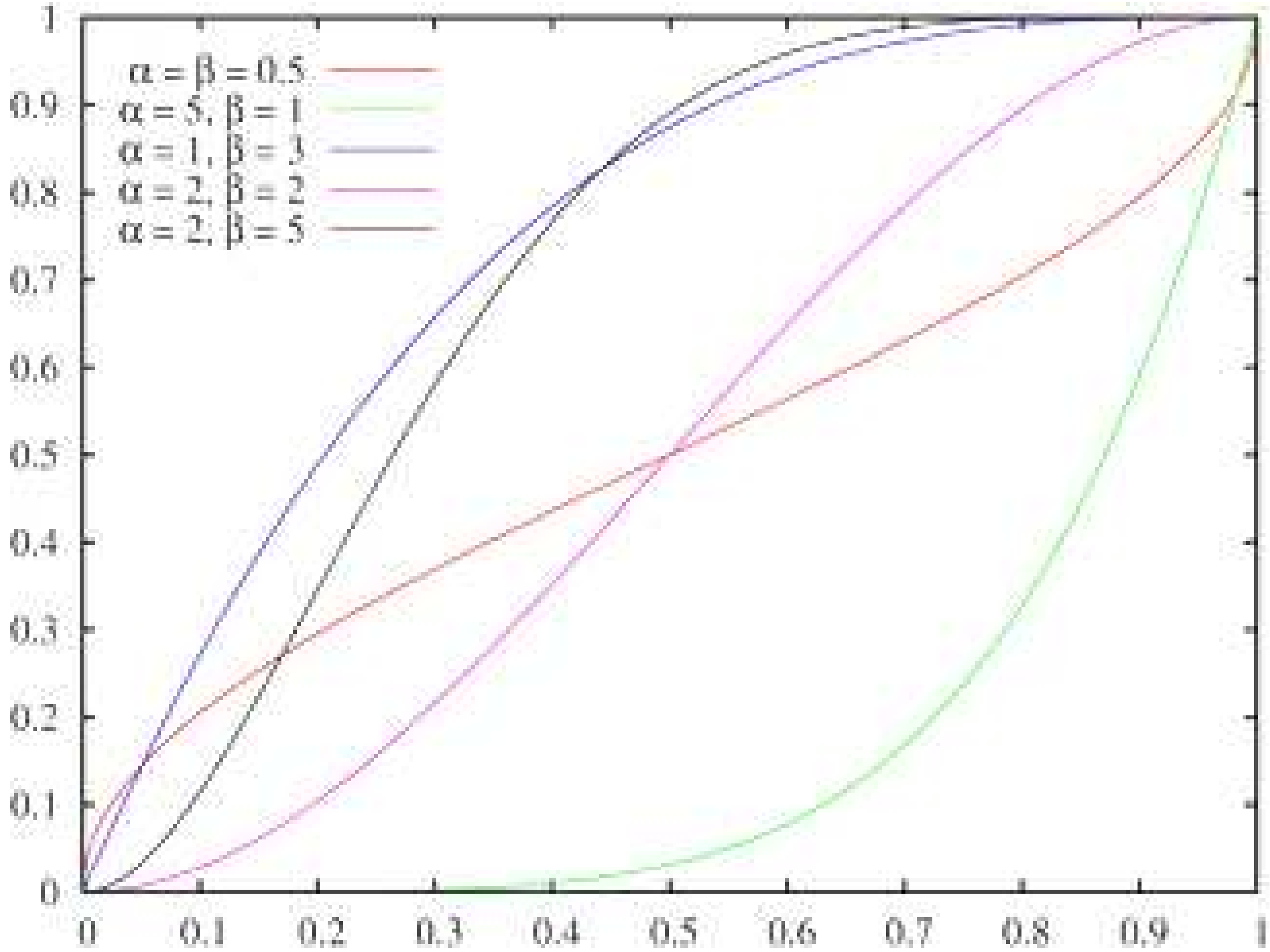
where $B_x(\alpha,\beta)$ is the incomplete beta function and $I_x(\alpha,\beta)$ is the regularized incomplete beta function.

- The general definition of a beta function is B(.,.)

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

for a>à, b>0

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


α = β = 0.5
α = 5, β = 1
α = 1, β = 3
α = 2, β = 2
α = 2, β = 5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Example

Rule of succession

- A classic application of the beta distribution is the rule of succession, introduced in the 18th century by Pierre-Simon Laplace in the course of treating the sunrise problem.
- It states that, given *s* successes in *n* conditionally independent Bernoulli trials with probability *p,* that *p* should be estimated as $\frac{s+1}{n+2}$.
- This estimate may be regarded as the expected value of the posterior distribution over *p,* namely Beta(*s* + 1, *n* − *s* + 1), which is given by Bayes' rule if one assumes a uniform prior over *p* (i.e., Beta(1, 1)) and then observes that *p* generated *s* successes in *n* trials.

Bayesian statistics

- Beta distributions are used extensively in Bayesian statistics, since beta distributions provide a family of conjugate prior distributions for binomial (including Bernoulli) and geometric distributions.
- The Beta(0,0) distribution is an improper prior and sometimes used to represent ignorance of parameter values.

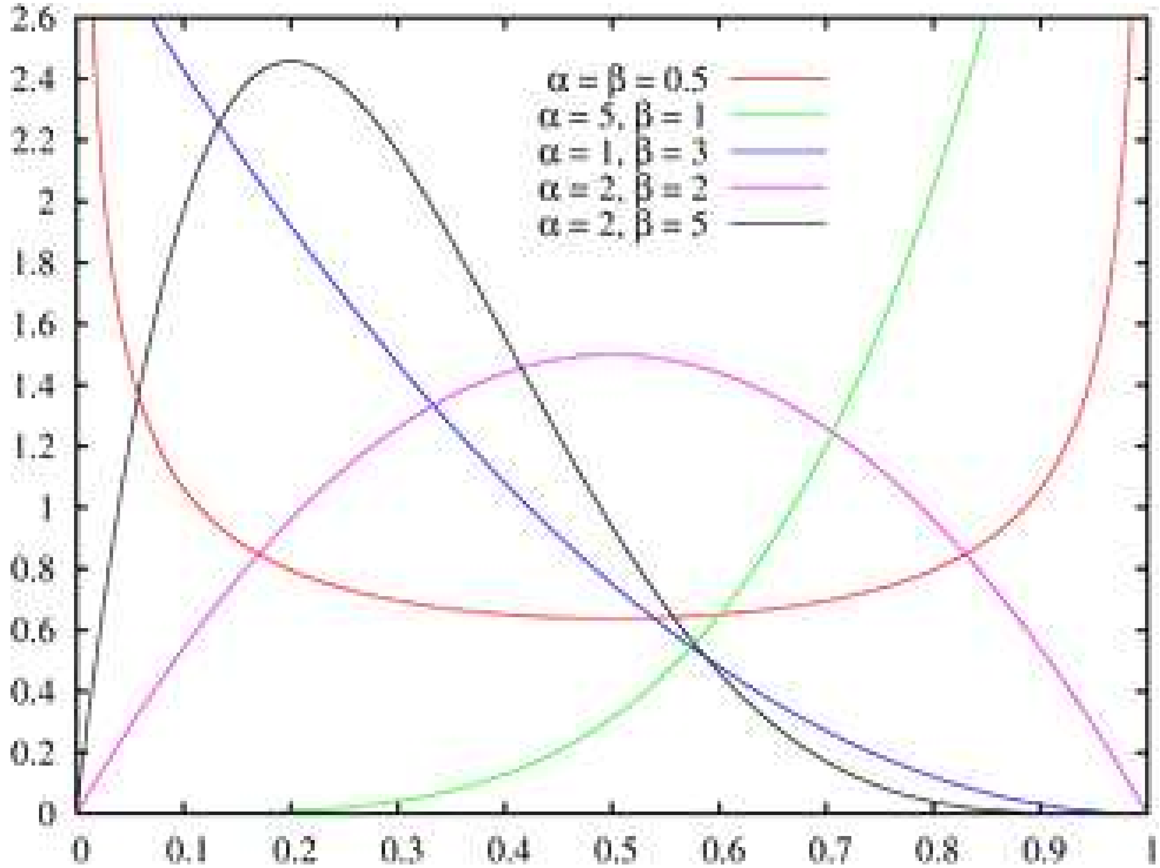
Task duration modeling

- The beta distribution can be used to model events which are constrained to take place within an interval defined by a minimum and maximum value. For this reason, the beta distribution is used extensively in critical path method (CPM) and other project management / control systems to describe the time to completion of a task.

- The beta density function can take on different shapes depending on the values of the two parameters:

  - $\alpha = 1,\ \beta = 1$ is the uniform [0,1] distribution
  - $\alpha < 1,\ \beta < 1$ is U-shaped (red plot)
  - $\alpha < 1,\ \beta \geq 1$ or $\alpha = 1,\ \beta > 1$ is strictly decreasing (blue plot)
    - $\alpha = 1,\ \beta > 2$ is strictly convex
    - $\alpha = 1,\ \beta = 2$ is a straight line
    - $\alpha = 1,\ 1 < \beta < 2$ is strictly concave
  - $\alpha = 1,\ \beta < 1$ or $\alpha > 1,\ \beta \leq 1$ is strictly increasing (green plot)
    - $\alpha > 2,\ \beta = 1$ is strictly convex
    - $\alpha = 2,\ \beta = 1$ is a straight line
    - $1 < \alpha < 2,\ \beta = 1$ is strictly concave
  - $\alpha > 1,\ \beta > 1$ is unimodal (purple & black plots)

Moreover, if α = β then the density function is symmetric about 1/2 (red & purple plots).



- The beta distribution reduces to the uniform distribution over (0,1) if α=β=1.

- A generalization of the complete beta function defined by

$$B(z; a, b) \equiv \int_0^z u^{a-1} (1-u)^{b-1} \, du,$$

sometimes also denoted $B_z(a, b)$. The so-called Chebyshev integral is given by

$$\int x^p (1-x)^q \, dx = B(x; 1+p, 1+q).$$

- The incomplete beta function is given in terms of hypergeometric functions by

$$B(z; a, b) = \frac{z^a}{a} \, {}_2F_1(a, 1-b; a+1; z)$$

$$= z^a \, \Gamma(a) \, {}_2\tilde{F}_1(a, 1-b; a+1; z).$$

- It is also given by the series

$$B(z; a, b) = z^a \sum_{n=0}^{\infty} \frac{(1-b)_n}{n!\,(a+n)} z^n.$$

- The incomplete beta function $B(z; a, b)$ reduces to the usual beta function $B(a, b)$ when $z = 1$,

$$B(1; a, b) = B(a, b).$$

Quantities of information

- Given two beta distributed random variables, X ~ Beta(α, β) and

Y ~ Beta(α', β'), the **information entropy** of *X* is

$$H(X) = \ln \mathrm{B}(\alpha, \beta) - (\alpha - 1)\psi(\alpha) - (\beta - 1)\psi(\beta) + (\alpha + \beta - 2)\psi(\alpha + \beta)$$

where ψ is the digamma function (= logarithmic derivative of the gamma function).

- The **cross entropy** can be shown to be

$$H(X,Y) = \ln \mathrm{B}(\alpha', \beta') - (\alpha' - 1)\psi(\alpha) - (\beta' - 1)\psi(\beta) + (\alpha' + \beta' - 2)\psi(\alpha + \beta).$$

- The Kullback-Leibler divergence can be considered as a “kind of” distance between two probability densities. It is however not a real distance measure, since it is not symmetric!

- It follows from the previous that the Kullback–Leibler divergence between the two beta distributions is

$$D_{\mathrm{KL}}(X,Y) = \ln \frac{\mathrm{B}(\alpha',\beta')}{\mathrm{B}(\alpha,\beta)} - (\alpha'-\alpha)\psi(\alpha) - (\beta'-\beta)\psi(\beta) + (\alpha'-\alpha+\beta'-\beta)\psi(\alpha+\beta).$$

4 Where discrete and continuous distributions meet

4.1 Approximations

Binomial by Poisson We defined the binomial discrete density function, with parameters n and p, as

$$\binom{n}{x} p^x(1-p)^{n-x} \qquad \text{for } x = 0, 1, \ldots, n.$$

If the parameter n approaches infinity and p approaches 0 in such a way that np remains constant, say equal to λ, then

$$\binom{n}{x} p^x(1-p)^{n-x} \to \frac{e^{-\lambda}\lambda^x}{x!}$$

for fixed integer x. The above follows immediately from the following consideration:

$$\binom{n}{x} p^x(1-p)^{n-x} = \frac{(n)_x}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$
$$= \frac{\lambda^x}{x!}\frac{(n)_x}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x} \to \frac{e^{-\lambda}\lambda^x}{x!},$$

since

$$\frac{(n)_x}{n^x} \to 1, \qquad \left(1 - \frac{\lambda}{n}\right)^{-x} \to 1, \qquad \text{and} \qquad \left(1 - \frac{\lambda}{n}\right)^{n} \to e^{-\lambda} \qquad \text{as} \quad n \to \infty.$$

Thus, for large n and small p the binomial probability $\binom{n}{x} p^x (1-p)^{n-x}$ can be approximated by the Poisson probability $e^{-np}(np)^x/x!$. The utility of this approximation is evident if one notes that the binomial probability involves two parameters and the Poisson only one.

Binomial and Poisson by normal

Theorem Let random variable X have a Poisson distribution with parameter λ; then for fixed $a < b$

$$P\left[a < \frac{X - \lambda}{\sqrt{\lambda}} < b\right]$$

$$= P[\lambda + a\sqrt{\lambda} < X < \lambda + b\sqrt{\lambda}] \to \Phi(b) - \Phi(a) \quad \text{as} \quad \lambda \to \infty.$$

Theorem **De Moivre–Laplace limit theorem** Let a random variable X have a binomial distribution with parameters n and p; then for fixed $a < b$

$$P\left[a \leq \frac{X - np}{\sqrt{npq}} \leq b\right] = P[np + a\sqrt{npq} \leq X \leq np + b\sqrt{npq}] \rightarrow$$

$$\Phi(b) - \Phi(a) \qquad \text{as } n \rightarrow \infty.$$

- This result is a special case of the central limit theorem
- It also gives a normal approximation for the binomial distribution with large n.

EXAMPLE Suppose that two fair dice are tossed 600 times. Let X denote the number of times a total of 7 occurs. Then X has a binomial distribution with parameters $n = 600$ and $p = \frac{1}{6}$. $\mathscr{E}[X] = 100$. Find $P[90 \leq X \leq 110]$.

$$P[90 \leq X \leq 110] = \sum_{j=90}^{110} \binom{600}{j} \left(\frac{1}{6}\right)^j \left(\frac{5}{6}\right)^{600-j},$$

a sum that is tedious to evaluate. Using the approximation given

$$P[90 \leq X \leq 110] \approx \Phi\left(\frac{110-100}{\sqrt{\frac{500}{6}}}\right) - \Phi\left(\frac{90-100}{\sqrt{\frac{500}{6}}}\right)$$

$$= \Phi(\sqrt{\tfrac{6}{5}}) - \Phi(-\sqrt{\tfrac{6}{5}}) \approx \Phi(1.095) - \Phi(-1.095) \approx .726.$$

////

4.2 Poisson and exponential relationships

When introducing the Poisson distribution, an experiment consisting of the counting of the number of happenings of a certain phenomenon in time was given special consideration. We argued that under certain conditions the count of the number of happenings in a fixed time interval was Poisson distributed with parameter, the mean, proportional to the length of the interval. Suppose now that one of these happenings has just occurred; what then is the distribution of the length of time, say X, that one will have to wait until the next happening? $P[X > t] = P[\text{no happenings in time interval of length } t] = e^{-\nu t}$, where ν is the *mean occurrence rate*; so

$$F_X(t) = P[X \le t] = 1 - P[X > t] = 1 - e^{-\nu t} \qquad \text{for } t > 0;$$

$\Rightarrow$ X follows an exponential distribution

On the other hand, it can be proved, under an independence assumption, that if the happenings are occurring in time in such a way that the distribution of the lengths of time between successive happenings is exponential, then the distribution of the number of happenings in a fixed time interval is Poisson distributed.

$\Rightarrow$ Exponential and Poisson are related

4.3 Deviations from the ideal world ?

Mixtures of distributions

A brief introduction to the concept of *contagious distributions* is given here. If $f_0(\cdot), f_1(\cdot), \ldots, f_n(\cdot), \ldots$ is a sequence of density functions which are either all discrete density functions or all probability density functions which may or may not depend on parameters, and $p_0, p_1, \ldots, p_n, \ldots$ is a sequence of parameters satisfying $p_i \geq 0$ and $\sum_{i=0}^{\infty} p_i = 1$, then $\sum_{i=0}^{\infty} p_i f_i(x)$ is a density function, which is sometimes called a *contagious* distribution or a *mixture*. For example, if $f_0(x) = \phi_{\mu_0, \sigma_0^2}(x)$ (a normal with mean μ_0 and variance σ_0^2) and $f_1(x) = \phi_{\mu_1, \sigma_1^2}(x)$, then

$$p_0 \phi_{\mu_0,\sigma_0^2}(x) + p_1 \; \phi_{\mu_1,\sigma_1^2}(x)$$

$$= (1-p)\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2}[(x-\mu_0)/\sigma_0]^2} + p\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}[(x-\mu_1)/\sigma_1]^2}$$

where $p_1 = p$ and $p_0 = 1 - p$, is a *mixture* of two normal densities.

Truncated distributions

A normal distribution that is truncated at 0 on the left and at 1 on the right is defined in density form as

$$f(x) = f(x; \mu, \sigma) = \frac{\phi_{\mu,\sigma^2}(x) I_{(0,1)}(x)}{\Phi_{\mu,\sigma^2}(1) - \Phi_{\mu,\sigma^2}(0)}.$$

This truncated normal distribution, like the beta distribution, assumes values between 0 and 1.

Truncation can be defined in general. If X is a random variable with density $f_X(\cdot)$ and cumulative distribution $F_X(\cdot)$, then the density of X truncated on the left at a and on the right at b is given by

$$\frac{f_X(x) I_{(a,b)}(x)}{F_X(b) - F_X(a)}.$$

ɔu need to divide by the difference of F_X evaluated in a and b?