Chapter 3
Examples of Mass Functions and Densities
Continuous Random Variables

Outline

Introduction

We will characterize a family of continuous random variables depending on the parameter θ, with their density

$$f_X(x|\theta) \approx \frac{1}{\Delta x} P_\theta\{x < X \leq x + \Delta x\}.$$

We will

- use the expression $Family(\theta)$ as shorthand for this family
- followed by the R command `family` and
- the state space S. The density is 0 outside of S.

Uniform Random Variables

$U(a, b)$ (R command `unif`) on $S = [a, b]$, $a < b$,

$$f_X(x|a, b) = \frac{1}{b - a}.$$

Independent $U(0, 1)$ are the most common choice for generating random numbers. Use the R command `runif(n)` to simulate `n` independent random numbers.

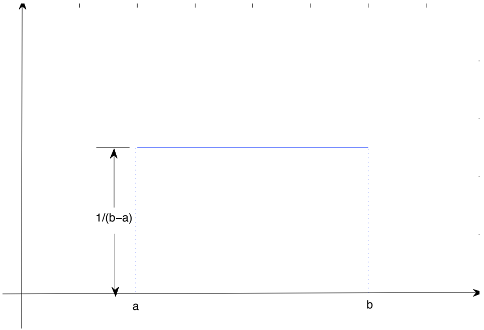


Figure: Uniform density on the interval $[a, b]$

Exponential Random Variables

$Exp(\lambda)$ (R command `exp`) on $S = [0, \infty)$

$$f_X(x|\lambda) = \lambda e^{-\lambda x}.$$

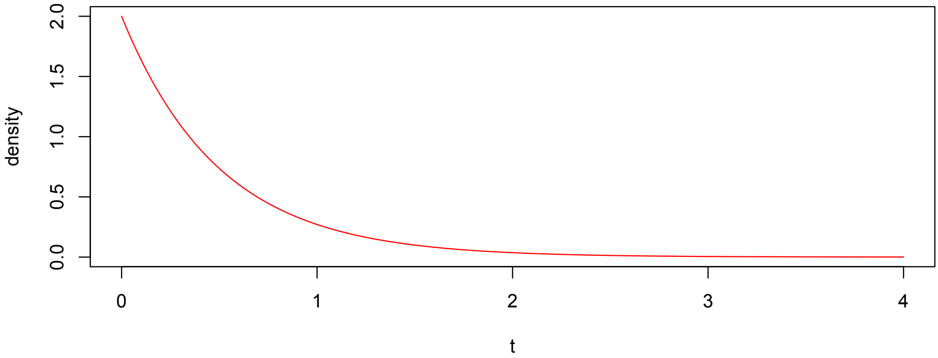


Figure: Exponential density with parameter $\lambda = 2$

Exponential Random Variables

Consider Bernoulli trials arriving at a rate of n trials per time unit.

- Let the probability of success p be small,
- nt the number of trials up to a given time t be large, and
- $\lambda = np$ be moderate in size.

Let T be the time of the first success. This random time exceeds a given time t if we begin with nt consecutive failures. The survival function

$$\bar{F}_T(t) = P\{T > t\} = (1-p)^{nt} = \left(1 - \frac{\lambda}{n}\right)^{nt} \approx e^{-\lambda t}.$$

The cumulative distribution function

$$F_T(t) = P\{T \leq t\} = 1 - P\{T > t\} \approx 1 - e^{-\lambda t}.$$

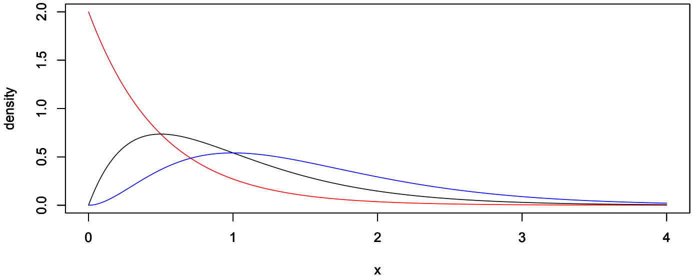
The density can be found by taking a derivative of $F_T(t)$.

Gamma Random Variables

$\Gamma(\alpha, \beta)$ (R command `gamma`) on $S = [0, \infty)$,

$$f_X(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

The gamma function $\Gamma(s) = \int_0^\infty x^s e^{-x} \frac{dx}{x}$, `gamma(s)` in R.



The gamma density with parameters $\beta = 2$ and $\alpha = 1$, 2, 3, $Gamma(\beta, 1)$ is $Exp(\beta)$

The Gamma Function

We can use integration by parts obtain a recursion relation for the gamma function,

$$\begin{aligned}\Gamma(s+1) = \int_0^\infty x^s e^{-x}\,dx &= -x^s e^{-x}\Big|_0^\infty - \int_0^\infty s x^{s-1}(-e^{-x})\,dx \\ &= s\int_0^\infty x^{s-1}e^{-x}\,dx = s\Gamma(s).\end{aligned}$$

$$\begin{aligned} u(x) &= x^s & v(x) &= -e^{-x} \\ u'(x) &= s x^{s-1} & v'(x) &= e^{-x}\end{aligned}$$

Exercise. $\Gamma(1) = 1$ and thus $\Gamma(n+1) = n!$

The Gamma Function

A change of variables will allow us to find the value of $\Gamma(1/2)$.

$$\begin{aligned}\Gamma(1/2) &= \int_0^\infty x^{-1/2} e^{-x}\,dx \\ &= \sqrt{2}\int_0^\infty e^{-z^2/2}\,dz = \sqrt{2}\cdot\frac{1}{2}\sqrt{2\pi} = \sqrt{\pi}\end{aligned}$$

$$z = \sqrt{2x}, \quad x = \frac{z^2}{2}, \quad dz = \frac{1}{\sqrt{2x}}dx, \quad x^{-1/2}dx = \sqrt{2}\,dz$$

Gamma Random Variables

Let n be a positive integer.

- a $\Gamma(n, \lambda)$ random variable can be seen as an approximation to the negative binomial random variable using the ideas that leads from the geometric random variable to the exponential.
- Alternatively, $\Gamma(n, \lambda)$ is the sum of n independent $Exp(\lambda)$ random variables.
  - This special case of the gamma distribution is sometimes called the Erlang distribution and was originally used in models for telephone traffic.

Beta Random Variables

Beta(α, β) (R command `beta`) on $S = [0, 1]$

$$f_X(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}.$$

Beta random variables appear in a variety of circumstances. One example is the order statistics. Beginning with n independent $U(0, 1)$ random variables and rank them

$$X_{(1)}, X_{(2)}, \dots, X_{(n)}$$

from smallest to largest. Then, the k-th order statistic $X_{(k)}$ is $Beta(k, n - k + 1)$.

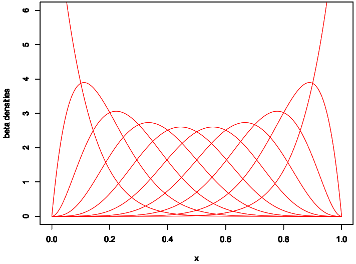


Figure: Densities of order statistics, $n = 10$

Normal Random Variables

$N(\mu,\sigma)$ (R command `norm`) on $S=\mathbb{R}$

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

Thus, a standard normal random variable is $N(0,1)$. Other normal random variables are linear transformations of Z, the standard normal. In particular, $X=\sigma Z+\mu$ has a $N(\mu,\sigma)$ distribution.

Exercise. If X is a normal random variable, then $Y=\exp X$ is called a log-normal random variable. Find $f_Y(y)$, the density for Y.

Answer. The transformation $g(x)=e^x$, its inverse $g^{-1}(y)=\ln y$ has derivative $\frac{d}{dy}g^{-1}(y)=\frac{1}{y}$. Thus,

$$f_Y(y)=f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|=\frac{1}{y\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln y-\mu)^2}{2\sigma^2}\right).$$

Logistic Random Variables

Logistic(μ, σ) (R command `logis`) on $S = \mathbb{R}$

The logistic random variable is generally introduced through its distribution function.

$$F_X(x) = \frac{1}{1 - \exp(-(x-\mu)/s)} = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x-\mu}{s}\right)$$

Exercise. Find the density function $f_X(x)$ for a logistic random variable.

Answer. We take the derivative

$$f_X(x) = \frac{\exp(-(x-\mu)/s))}{s(1 - \exp(-(x-\mu)/s))^2}$$

The final examples, each of them derived from normal random variables, will be used in hypothesis testing. In practice, probabilities and quantiles are generally computed using software rather than working with the densities explicitly

Student's t Random Variables

$t_\nu(\mu, \sigma)$ (R command `t`) on $S = \mathbb{R}$

$$f_X(x|\nu, \mu, \sigma) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)\sigma}\left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}.$$

The value ν is also called the number of degrees of freedom. If $\bar{Z}$ is the sample mean of n standard normal random variables and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2 \quad \text{is the sample variance}$$

$$\text{then,} \quad T = \frac{\bar{Z}}{S/\sqrt{n}}.$$

has a $t_{n-1}(0, 1)$ distribution.

χ^2 and F Random Variables

χ^2_ν (R command `chisq`) on $S = [0, \infty)$

$$f_X(x|\nu) = \frac{x^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)} e^{-x/2}.$$

The value ν is called the number of degrees of freedom. For ν a positive integer, χ^2_ν is the sum of the the squares of ν standard normal random variables.

F_{ν_1,ν_2} (R command `f`) on $S = [0, \infty)$

$$f_X(x|\nu_1, \nu_2) = \frac{\Gamma((\nu_1+\nu_2)/2)\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} x^{\nu_1/2-1}(\nu_2+\nu_1 x)^{-(\nu_1+\nu_2)/2}.$$

The F distribution is used in analysis of variance tests. F is the ratio of independent $\chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$ random variables.

Summary

random variable	R	parameters	mean	variance	characteristic function
beta	`beta`	α, β	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$F_{1,1}(a, b; \frac{i\theta}{2\pi})$
chi-squared	`chisq`	ν	ν	2ν	$\frac{1}{(1-2i\theta)^{\nu/2}}$
exponential	`exp`	λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{i\lambda}{\theta+i\lambda}$
log-normal	`lnorm`	μ, σ	$\exp(\mu + \sigma^2/2)$	$(e^{\sigma^2} - 1)\exp(2\mu + \sigma^2)$	
F	`f`	ν_1, ν_2	$\frac{\nu_2}{\nu_2-2}, \nu_2 > 2$	$2\nu_2^2 \frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)(\nu_2-2)^2}$	
logistic	`logis`	μ, s	μ	$s^2\pi^2/3$	$\pi e^{it\mu} st/\sinh(\pi st)$
gamma	`gamma`	α, β	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\left(\frac{i\beta}{\theta+i\beta}\right)^\alpha$
normal	`norm`	μ, σ^2	μ	σ^2	$\exp(i\mu\theta - \frac{1}{2}\sigma^2\theta^2)$
Pareto	`pareto`	α, β	$\frac{\alpha\beta}{\beta-1}, (\beta > 1)$	$\frac{\alpha^2\beta}{(\beta-1)^2(\beta-2)}, (\beta > 2)$	
t	`t`	ν, a, σ	$a, (\nu > 1)$	$\sigma^2 \frac{a}{a-2}, (\nu > 2)$	
uniform	`unif`	a, b	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$-i\frac{\exp(i\theta b)-\exp(i\theta a)}{\theta(b-a)}$

R Commands

R can compute a variety of values for many families of distributions.

- `dfamily(x, parameters)` is the mass function (for discrete random variables) or probability density (for continuous random variables) of `family` evaluated at x.
- `qfamily(p, parameters)` returns x satisfying $P\{X \leq x\} = p$, the p-th quantile where X has the given distribution,
- `pfamily(x, parameters)` returns $P\{X \leq x\}$ where X has the given distribution.
- `rfamily(n, parameters)` generates n random variables having the given distribution.

R Commands

Exercise.

- Find the quintiles of a gamma random variable with $\alpha = 4$ and $\beta = 5$
```
quintile<-seq(0,0.8,0.2); value<-qgamma(quintile,4,5);
data.frame(quintile,value)
```
  Now draw the density function and show the quintiles.
```
curve(dgamma(x,4,5),0,3);lines(value,dgamma(value,4,5),type="h",lty=2)
```
- Simulate 1000 independent observations of a $Beta(2,5)$ random variable and gives the quartiles of the simulation.
```
x<-rbeta(1000,2,5); quantile(x)
```
  Compare this to the distribution values.
```
p<-0:4/4; qbeta(p,2,5)
```