Topic 9

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.

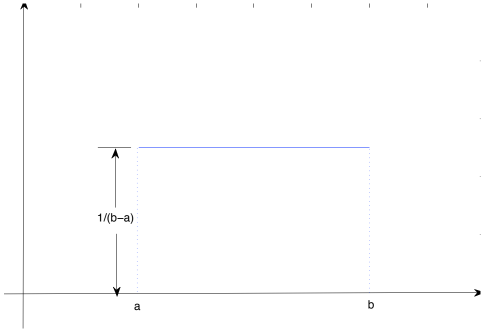


Figure: Uniform density on the interval $[a,b]$

Exponential Random Variables

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

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

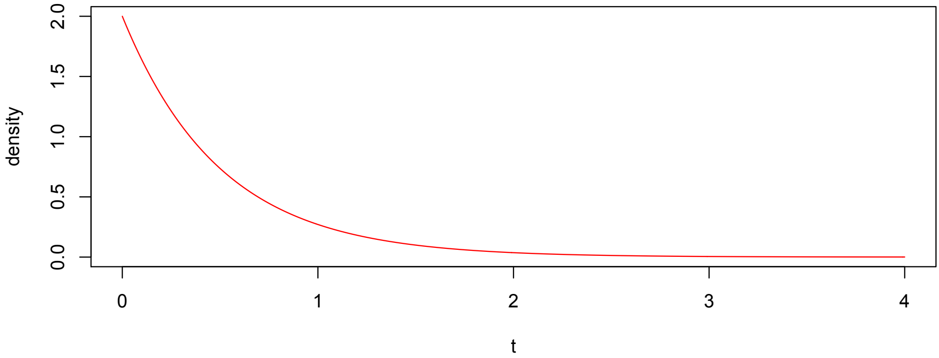


Figure: Exponential density with parameter $\lambda = 2$

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

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

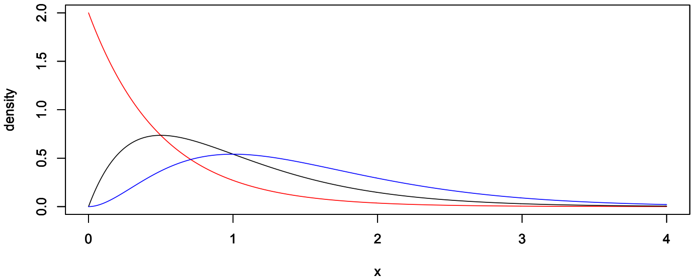
The density can be found by taking a derivative of $F_T(t)$.

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 $Exp(\beta)$

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

$Beta(\alpha, \beta)$ (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

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

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

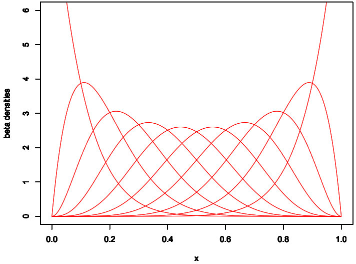


Figure: Densities of order statistics, $n = 10$

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.

The final examples, each of them derived from normal random variables, will be used in hypothesis testing. In practice, probabilities 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

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

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

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.

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.

1. Find $P\{X = x\}$ for $x = 0, 1, 2, 3, 4, 5$ for a $Bin(5, 3/7)$ random variable.
2. Find the first and third quartiles as well as the median for a $Beta(3, 3)$ random variables.
3. Find $P\{X \leq x\}$ for $x = 0, 1, 2, 3$ for a χ^2_2 random variable.
4. Simulate 80 $Pois(5)$ random variable. Find the mean and variance of these simulated values.