Distributions cheatsheet

Gavin Band, WHG GMS Programme 2021

B = beta or multivariate beta function Γ = gamma function

$\binom{N}{k}$ = N choose k Red terms are those that do not involve x

See also: the distribution zoo

See also: A longer list of probability distributions

Type	Domain	Name	Parameters	Mass / density	Distribution of…	R pdf function	Applications
Continuous unbounded	$x \in \mathbb{R}$	Gaussian or normal	Mean μ Variance σ^2	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$	Sum of independent scalars (e.g. errors)	`dnorm()`	Ubiquitous
	$x \in \mathbb{R}^d$	Multivariate normal	Mean vector μ Covariance matrix Σ	$\frac{1}{\sqrt{2\pi\lvert\Sigma\rvert}} e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}$	Sum of independent vectors	`dmvnorm()` (mvtnorm package)	Ubiquitous
	$x \in \mathbb{R}$	Laplace or double exponential	Location μ Scale $b > 0$	$\frac{1}{2b} e^{-\frac{\lvert x-\mu\rvert}{b}}$		Use `dexp()`	Regularised (Lasso) regression
Continuous bounded	$x \in [0,1]$ or $(0,1)$	Beta	α (= "shape1") β (= "shape2")	$\frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	Order statistics for uniform values	`dbeta()`	Prior for binomial-distributed variables
	$x \in \Delta_{d-1}$ (d-1)-dimensional simplex	Dirichlet	"shape" parameters $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{R}^d$	$\frac{1}{B(\alpha)} \prod_i x_i^{\alpha_i - 1}$			Prior for multinomial-distribution variables
Continuous positive	$[0, \infty)$	Exponential	Rate λ	$\lambda e^{-\lambda x}$	Time between events occuring at fixed rate	`dexp()`	Modelling waiting times between rare events
	$[0, \infty)$ or $(0, \infty)$	Chi-squared	Degrees of freedom k	$\frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$	Sum of squared Gaussian variables	`dchisq()`	Likelihood ratio test
Discrete	$x \in \{0,1\}$	Bernoulli	Success probability p	$p^x(1-p)^{1-x}$	Coin flip	`dbinom()`	Logistic regression
	$x \in \{0, \cdots, n\}$	Binomial	Number of trials n Success probability p	$\binom{n}{x} p^x (1-p)^{n-x}$	Coin flips / sample with replacement	`dbinom()`	Logistic regression
	$x \in 0, 1, \cdots$	Geometric	Success probability p	$(1-p)^x p$	Number of Bernoulli trials before 1st success	`dgeom()`	
	$(x_i) \in \{0, \cdots, n\}^d$ $\sum x_i = n$	Multinomial or categorical	Outcome probabilities $p = (p_1, p_2, \cdots, p_d)$	$\frac{n!}{x_1! \cdots x_d!} p_1^{x_1} \cdots p_d^{x_d}$	Possible outcomes	`dmultinom()`	Multinomial logistic regression
	$x \in \{0, \cdots, n\}$	Hypergeometric	Population size N Total no. of successes K No. of draws n	$\frac{\binom{K}{x}\binom{N-K}{n-k}}{\binom{N}{n}}$	Sample without replacement	`dhyper()`	Fisher's exact test
Count	$x \in 0, 1, \cdots$	Poisson	Rate λ	$e^{-\lambda} \frac{\lambda^x}{x!}$	Number of events occuring at fixed rate in a unit of time	`dpois()`	Modelling rare events (e.g. sequence reads along genome)