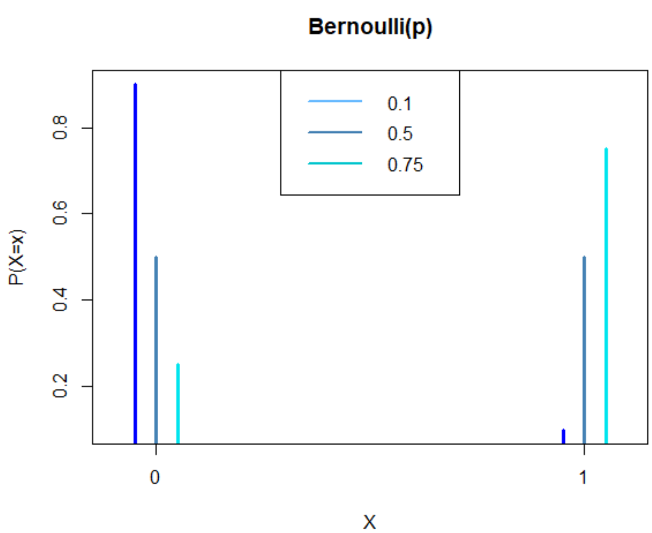
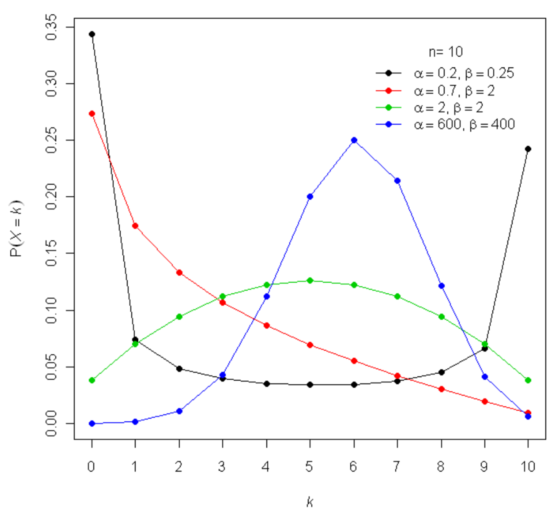
Bernoulli(p)

pmf (& support)	$f(x\|p) = p^x(1-p)^{1-x}$ for $x = 0, 1$
parameter space	$0 < p < 1$ where $p = P(X = 1) = P(\text{success})$ (Note: technically $0 \leq p \leq 1$, but sometimes the edge cases cause problems.)
mean & variance	$E(X) = p$ $\text{Var}(X) = p(1-p)$
mgf	$M_X(t) = pe^t + (1-p)$
graph	
shape	Since X can only be two values, there isn't much of a defined shape.
common uses	Usually used to model a binary "experiment" where the outcome is either a 1 (success) or 0 (failure), for example: win/lose; heads/tails; live/die, etc.

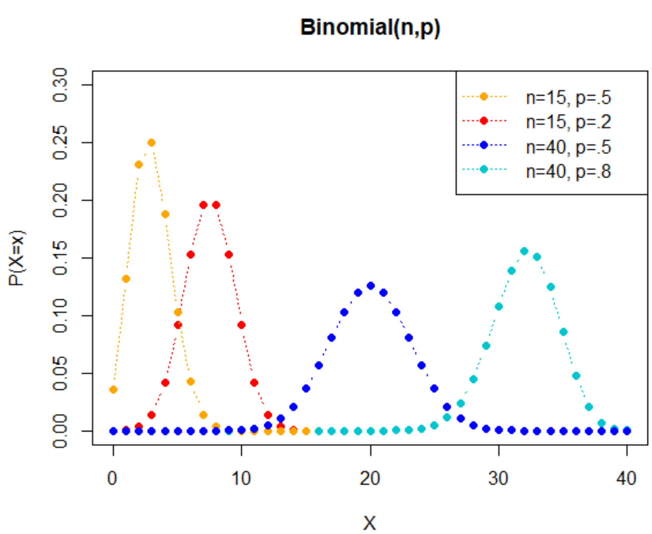
R functions	`dbinom(x, 1, p)`
	`rbinom(n, 1, p)`
	(Note `qbinom` and `pbinom` also exist but aren't very useful for Bernoulli)

special cases & relationships	- Special case of the binomial distribution when $n = 1$ - Sum of n Bernoulli(p) trials is binomial(n,p) - Number of Bernouli(p) *trials* until r^{th} success is Negative binomial(p, r)
Random Generation	If runif(0,1)$< p$, $X = 1$, else $X = 0$.

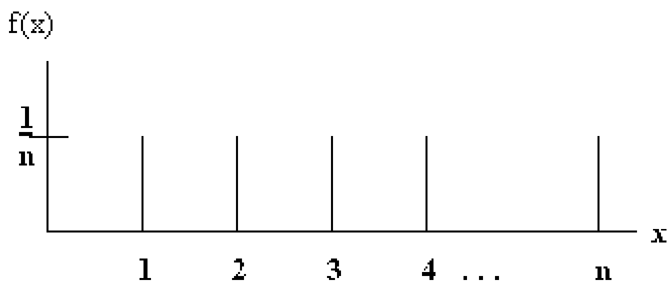
Beta-Binomial(α, β, n)

pmf (& support)	$f(x\|\alpha,\beta,n) = \binom{n}{x}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \binom{n}{x}\frac{B(x+\alpha,n-x+\beta)}{B(\alpha,\beta)} \quad x = 0, 1, \ldots n$ Where $B(,)$ is the beta function
parameter space	$\alpha \geq 0$ and $\beta \geq 0$ (shape parameters); n (number of trials) is a positive integer
mean & variance	$E(X) = \frac{n\alpha}{\alpha+\beta}$ $\qquad \text{Var}(X) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$
mgf	The form of the mgf is complicated and unhelpful
graph	
shape	If $\alpha = \beta$, the distribution will be symmetric; if $\alpha < \beta$, the distribution will be right skewed; if $\alpha > \beta$, the distribution will be left skewed; if both $\alpha < 1$ and $\beta < 1$, the distribution will be sort of "U" shaped. As n gets larger, distribution will look more smooth
common uses	Used as a predictive or marginal distribution in Bayesian statistics with a beta-binomial model. Can be interpreted as a weighted binomial distribution over different values of p (which has a beta distribution)
R functions	R Does not have built in functions, but the pmf can be evaluated directly through beta functions (see pmf above). See "random generation' below for generating random draws.
special cases & relationships	- When $\alpha = 1$ and $\beta = 1$, it reduces to a Discrete Uniform - If we let $p = \frac{\alpha}{\alpha+\beta}$ and as $n \to \infty$, this approaches a binomial
Random Generation	Let $X \sim Binomial(n, P)$ and $P \sim Beta(\alpha, \beta)$. Draw Ps, use those to get X.

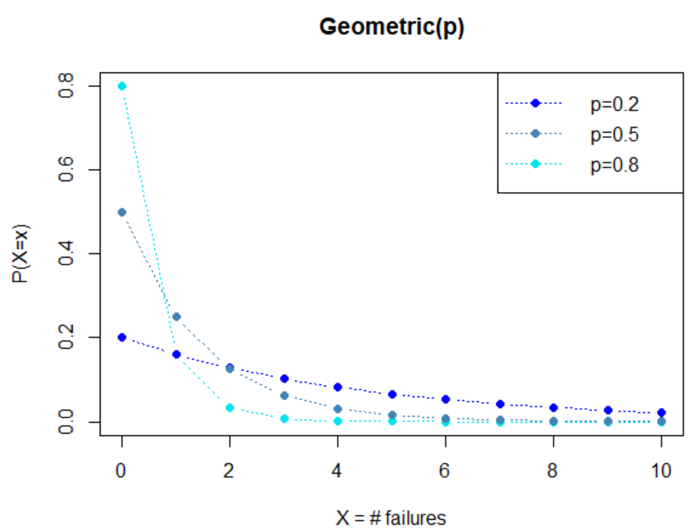
Binomial(n, p)

pmf (& support)	$f(x\|n,p) = \binom{n}{x} p^x (1-p)^{1-x}$ for $x = 0, 1, ..., n$
parameter space	$0 < p < 1$, where $p = P(\text{success})$; n (number of trials) is a positive integer Note: n is usually considered known
mean & variance	$E(X) = np$ $\text{Var}(X) = np(1-p)$
mgf	$M_X(t) = (pe^t + (1-p))^n$
graph	
shape	If p is close to 0 or 1, graph is skewed. As n gets larger or if p is close 0.5, it looks more normal.
common uses	Modeling the number of successes in a fixed number (n) trials, sampling with replacement. Ex. Number of heads in 10 coin flips, number of 6s after rolling n dice.
R functions	`dbinom(x, n, p)` `pbinom(x, n, p)` `qbinom(q, n, p)` `rbinom(nDraws,n, p)`
special cases & relationships	- When $n = 1$, it's a Bernoulli - As $n \to \infty$, it approaches a normal - Can approximate a Poisson if $p = \lambda n$ and n is large, - Can approximate Hypergeometric(M, N, k) $N \to \infty$
Random Generation	Let $Y_i \sim Bernoulli(p)$, then $X = \sum_{i=1}^{n} Y_i$ (sum of n independent Bernoulli varibules).

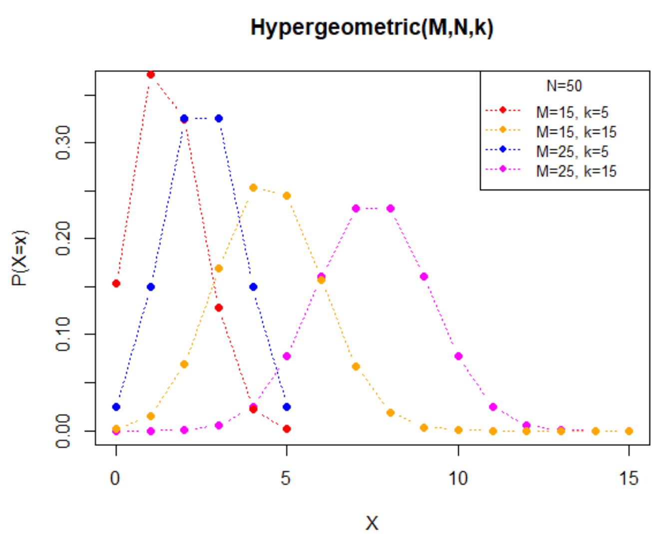
Discrete Uniform(N)

pmf (& support)	$f(x\|N) = 1/N$ for x = 1,2,...,N
parameter space	N is a positive integer
mean & variance	$E(X) = \frac{N+1}{2}$ $\mathrm{Var}(X) = \frac{N^2-1}{12}$
mgf	$M_X(t) = \sum_{i=1}^{N} e^{it}$
graph	
shape	Graph has N "tick marks" of all equal height at integers 1 to N.
common uses	Discrete outcomes that are all equally likely. Ex. Rolling an N sided die.
R functions	No default R functions. Though, these aren't hard to calculate since the pmf is uniform.
special cases & relationships	- Special case of Beta-Binomial(α, β) when $\alpha = 1$ and $\beta = 1$ - Can also be parameterized using support a, a+1, ..., b-1, b
Random Generation	$X = \lceil runif(0, N) \rceil$ Where the function $\lceil . \rceil$ always rounds up.

Geometric(p)

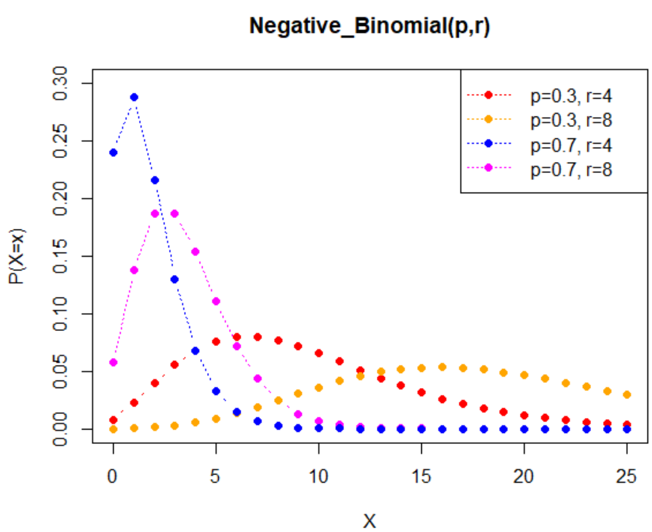
<table>
<tr><td>Parameterization</td><td>X = Number of Failures</td><td>X = Number of Trials</td></tr>
<tr><td>pmf (& support)</td><td>$f(x|p) = p(1-p)^x$ for x =0,1,...</td><td>$f(x|p) = p(1-p)^{x-1}$ for x =1,2,...</td></tr>
<tr><td>parameter space</td><td colspan="2">$0 < p < 1$ where $p = P(\text{success})$</td></tr>
<tr><td>mean & variance</td><td>$E(X) = \frac{1-p}{p}$ $\quad$ $\text{Var}(X) = \frac{1-p}{p^2}$</td><td>$E(X) = \frac{1}{p}$ $\quad$ $\text{Var}(X) = \frac{1-p}{p^2}$</td></tr>
<tr><td>mgf</td><td>$M_X(t) = \frac{p}{1-(1-p)e^t}$ $\quad$ $t < -ln(1-p)$</td><td>$M_X(t) = \frac{pe^t}{1-(1-p)e^t}$ $\quad$ $t < -ln(1-p)$</td></tr>
<tr><td>graph</td><td colspan="2">
</td></tr>
<tr><td>shape</td><td colspan="2">Looks like a discrete exponential. Higher values of p have more mass near X=1, lower values of p have more disperse mass.</td></tr>
<tr><td>common uses</td><td colspan="2">Used to model number of trials until first success.
Ex. Number of coin flips until first heads.</td></tr>
<tr><td>R functions</td><td colspan="2">Note: These are only for the $X = 0, 1, ...$ (number of failures) parameterization
<code>dgeom(x, p)</code>   <code>pgeom(x, p)</code>
<code>qgeom(q, p)</code>   <code>rgeom(n, p)</code></td></tr>
<tr><td>special cases & relationships</td><td colspan="2">- Special case of the negative binomial distribution when $r = 1$
- Sum of n $Geometric(p)$ trials is $NegativeBinomial(n, p)$
- Distribution has memory-less property
- If $X \sim Exponential(1)$, $\lfloor X \rfloor \sim Geometric(1 - \frac{1}{e})$</td></tr>
<tr><td>Random Generation</td><td colspan="2">Run Bernoulli trials until you get a 1. X = number of failures (0's) or trials.</td></tr>
</table>

Hypergeometric(N, M, k)

pmf (& support)	$f(x\|N,M,k) = \frac{\binom{M}{x}\binom{N-M}{k-x}}{\binom{N}{k}}$ for $x = max(0, k+M-N), 1, 2, .., min(k, M)$
parameter space	N = Population size, where $N > 0$ M = Number of successes in the population, where $0 < M \leq N$ k = sample size, where $0 < k \leq N$
mean & variance	$E(X) = \frac{kM}{N}$ $\text{Var}(X) = \frac{kM}{N}\frac{N-M}{N}\frac{N-k}{N-1}$
mgf	Too complicated to be useful
graph	
shape	If both M and k are close to 0 or N, graph is skewed. As N gets larger, graph appears more normal.
common uses	Modeling the number of successes in a finite population and sampling without replacement. Ex. Number of aces drawn in a 13 card hand, number of defective parts in a sample, accuracy of a voting sample.
R functions	`dhyper(x, M, N, k)` `phyper(x, M, N, k)` `qhyper(p, M, N, k)` `rhyper(n, M, N, k)`
special cases & relationships	- Let $p = \frac{M}{N}, n = k$ and $N \to \infty$, this approaches a binomial(n,p). - If $k = 1$ this is bernoulli(M/N)
Random Generation	Generate a population of size N with M successes. Sample k elements without replacement. X = number of successes in sample.

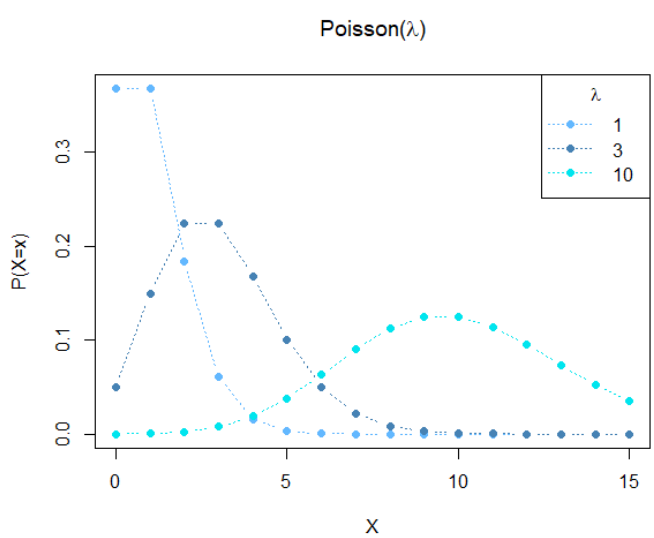
Negative Binomial(r, p)

<table>
<tr><td>Parameterization</td><td>X = Number of Failures</td><td>X = Number of Trials</td></tr>
<tr><td>pmf (& support)</td><td>$f(x|r,p) = \binom{r+x-1}{x} p^r (1-p)^x$ for x =0,1,...</td><td>$f(x|r,p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$ for x =r,r+1,...</td></tr>
<tr><td>parameter space</td><td colspan="2">$0 < p < 1$ where $p = P(\text{success})$; Experiment stops after r successes where r is a positive integer</td></tr>
<tr><td>mean & variance</td><td>$E(X) = \frac{r(1-p)}{p}$ $\quad$ $\text{Var}(X) = \frac{r(1-p)}{p^2}$</td><td>$E(X) = \frac{r}{p}$ $\quad$ $\text{Var}(X) = \frac{r(1-p)}{p^2}$</td></tr>
<tr><td>mgf</td><td>$M_X(t) = (\frac{p}{1-(1-p)e^t})^r \quad t < -ln(1-p)$</td><td>$M_X(t) = (\frac{pe^t}{1-(1-p)e^t})^r \quad t < -ln(1-p)$</td></tr>
<tr><td>graph</td><td colspan="2">

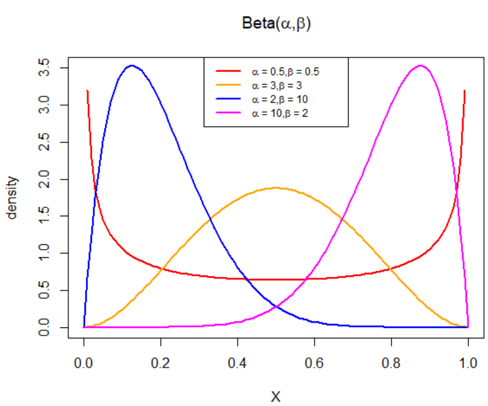


</td></tr>
<tr><td>shape</td><td colspan="2">Higher values of r make it look more normal. Higher values of p have more mass near X=0, lower values of p have more disperse mass. Graph is always skewed right.</td></tr>
<tr><td>common uses</td><td colspan="2">Used in experiments that stop after finding r successes.
Ex. Number of coin flips until $r = 3$ heads; if people decline surveys with prob p, how many people should we ask to get r participants?</td></tr>
<tr><td>R functions</td><td colspan="2">Note: These are only for the $X = 0, 1, ...$ (number of failures) parameterization
<code>dnbinom(x, r, p)</code> <code>pnbinom(x, r, p)</code>
<code>qnbinom(x, r, p)</code> <code>rnbinom(x, r, p)</code></td></tr>
<tr><td>special cases & relationships</td><td colspan="2">- Result of a sum of r independent geometric(p) variables (for matching parameterizations)
- If $r = 1$, it's a geometric.
- Can approximate Poisson if $\lambda = r(1-p)$ and as $r \to \infty$.</td></tr>
<tr><td>Random Generation</td><td colspan="2">Draw $U_1, U_2, ...U_r \sim$ unif(0,1) variables. X $= \lfloor \sum_{i=1}^{i=r} ln(U_i)/ln(1-p) \rfloor$ where $\lfloor . \rfloor$ means round down.</td></tr>
</table>

Poisson(λ)

pmf (& support)	$f(x\|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, \ldots$
parameter space	$\lambda > 0$, λ is interpreted as a rate parameter over a given interval of time or space.
mean & variance	$E(X) = \lambda$ $\text{Var}(X) = \lambda$
mgf	$M_X(t) = e^{\lambda(e^t - 1)}$
graph	
shape	If λ is low, graph looks like an exponential. As λ gets larger, the graph is more disperse and looks more normal.
common uses	Counting occurrences over a fixed interval of time or space. Ex. Number of scratches on a car surface, number of asteroids during a time interval.
R functions	`dpois(x, lambda)` `ppois(x, lambda)` `qpois(q, lambda)` `rpois(n,lambda)`
special cases & relationships	Can be approximated as a Normal(λ, λ) if λ is large - Sum of Poisson variables is also Poisson
Random Generation	Draw $U_1, U_2, .. \sim$ unif(0,1), $X = j - 1$ where j is the lowest index such that $\prod_{i=1}^{j} U_i < e^{-\lambda}$

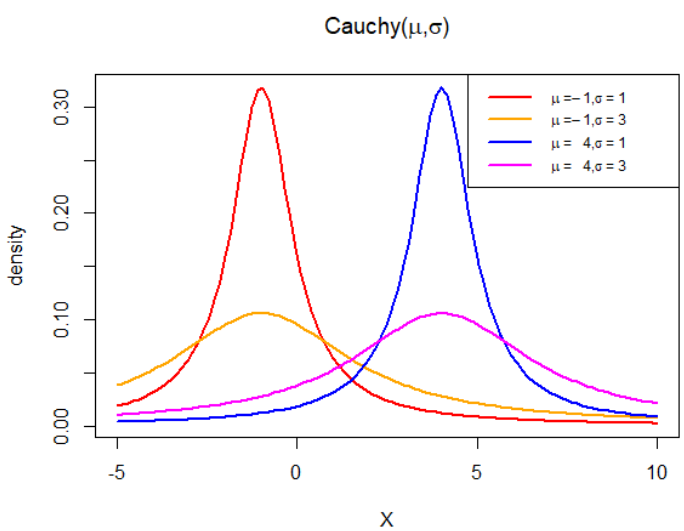
Beta(α, β)

pdf (& support)	$f(x\|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad \text{for } 0 \leq x \leq 1$
parameter space	$\alpha \geq 0$ and $\beta \geq 0$, both are called "shape" parameters
mean & variance	$E(X) = \frac{\alpha}{\alpha+\beta} \qquad \text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
mgf	The form of the mgf is complicated and unhelpful. However, raw moments can be calculated: $E(X^n) = \frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}$
graph	
shape	Lot of possible shapes: If $\alpha = \beta$, the distribution will be symmetric; if $\alpha < \beta$, the distribution will be right skewed; if $\alpha > \beta$, the distribution will be left skewed; if both $\alpha < 1$ and $\beta < 1$, the distribution will be sort of "U" shaped.
common uses	Commonly used in Bayesian inference as a prior distribution for parameters that are between 0 and 1 (such as for the Bernoulli, binomial, geometric, and negative binomial distributions).

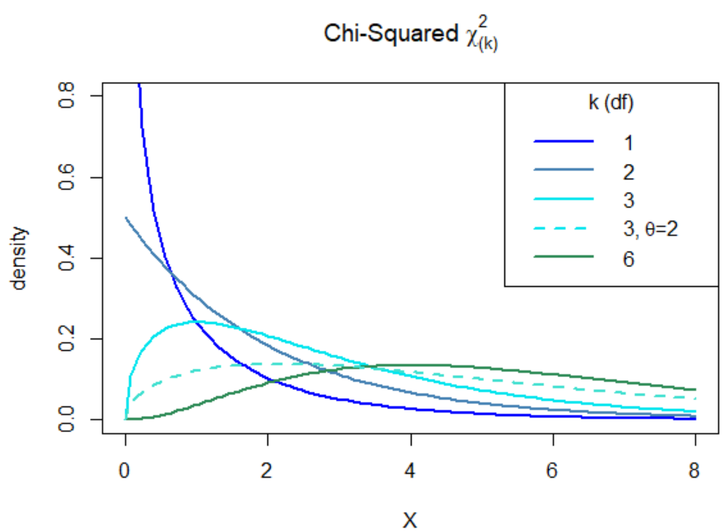
R functions	`dbeta(x, alpha, beta)` `pbeta(x, alpha, beta)`
	`qbeta(p, alpha, beta)` `rbeta(n, alpha, beta)`
	Note: `R` documentation refers to α as `shape1` and β as `shape2`

special cases & relationships	When $\alpha = 1$ and $\beta = 1$ reduces to uniform(0,1) A gamma over the sum of two independent gammas is a beta. Same for Chi-Squared and F. If $\alpha = \beta \to \infty$ then the beta converges to a normal
Random Generation	Let $Y \sim Gamma(\alpha, 1)$ and $Z \sim Gamma(\beta, 1)$. X $= \frac{Y}{Y+Z}$
Other notes (Beta function):	$\beta(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

Cauchy(μ, σ)

pdf (& support)	$f(x\|\lambda) = \frac{1}{\sigma\pi(1+(\frac{x-\mu}{\sigma})^2)}$ for $-\infty < x < \infty$
parameter space	$-\infty < \mu < \infty$ (location); $\sigma > 0$ (scale)
mean & variance	$E(X) =$ Undefined $\quad$ $\text{Var}(X) =$ Undefined
mgf	Does not Exist
graph	
shape	The graph looks like a normal distribution with very heavy tails. μ affects where it is centered, σ scales it.
common uses	Used for heavy-tailed data or as a test distribution with an undefined mean/variance.
R functions	`dcauchy(x,`μ,σ`)` `pcauchy(x,`μ,σ`)` `qcauchy(p,`μ,σ`)` `rcauchy(n,`μ,σ`)`
special cases & relationships	- Special case of the t-distribution with 1 df. - Standard cauchy can be obtained by taking the ratio of 2 standard normals. - If $X \sim$ Cauchy Sampling distrubution of $\overline{x}$ is the same distribution as X.
Random Generation	Draw $U \sim$ unif(0,1), $X = \mu + \sigma \tan(\pi(U - 1/2)$ (Inverse CDF)

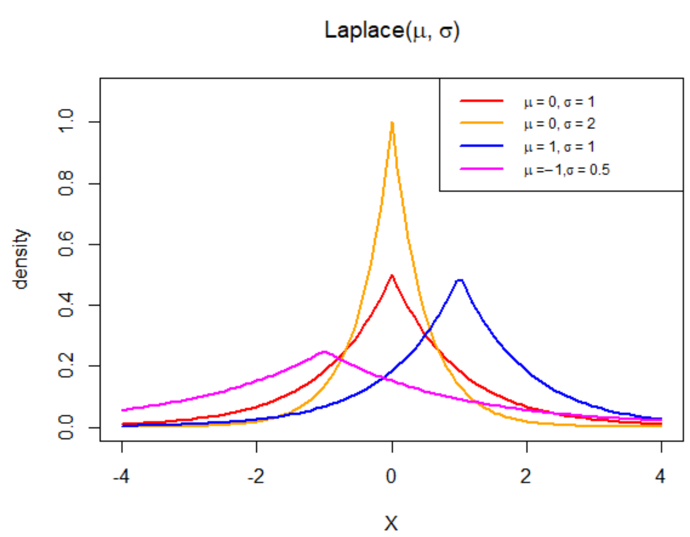
Chi-Squared(k)

pdf (& support)	$f(x\|k) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{k/2-1} e^{-x/2}$ for $x \geq 0$
parameter space	k (degrees of freedom) is a positive integer Also, non-centrality parameter (ncp) is often denoted $\theta \geq 0$
mean & variance	$E(X) = k + 2\theta$ $\quad$ $\text{Var}(X) = 2k + 8\theta$ $\quad$ (Using 1/2 parameterization of θ)
mgf	$M_X(t) = (\frac{1}{1-2t})^{k/2}$ $\quad t < \frac{1}{2}$; $\quad$ Non-Central χ^2: $(\frac{1}{1-2t})^{k/2} exp(\frac{2t\theta}{1-2t})$
graph	
shape	If $k < 2$ it has an asymptote at 0, if $k = 2$ it's exponential(1/2) $k > 2$ the density is shifted away from zero and it is right skewed. The ncp θ shifts density away from zero.
common uses	Used to test uniformity/independence of categorical variables, used in goodness of fit tests, derivations of the F and t-distribution, and in construction of confidence intervals.

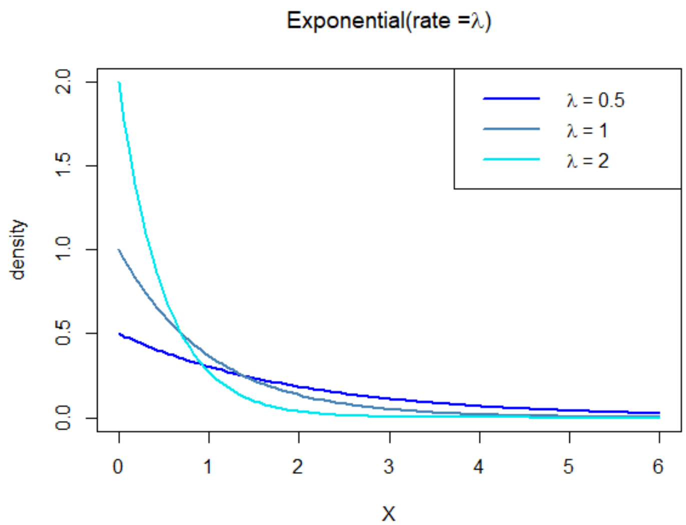
R functions	`dchisq(x,k, ncp)` `pchisq(x,k, ncp)`
	`qchisq(p,k, ncp)` `rchisq(n,k, ncp)`
	Where ncp is the non-centrality parameter, which is $\sum_i \mu_i^2$ in R

special cases & relationships	- Special case of the gamma when $\alpha = k/2$ and $\beta = 2$ - $\chi^2_{(k)}$ distribution is the sum of k standard normal variables. - If $Y \sim$ Normal, then $\frac{(n-p)s^2}{\sigma^2} \sim \chi^2_{n-p}$ - If $Y_1, ...Y_k \sim$ Normal$(\mu_i, 1)$, then $\sum_{i=1}^k Y_i^2 \sim \chi^2(k, ncp)$ where $ncp = \frac{1}{2}\sum_{i=1}^n u_i^2$
Random Generation	Draw $Y_1, ...Y_k \sim$ Normal(0,1), then $X = \sum_{i=1}^k Y_i^2$. Chi-squared with ncp can be drawn by changing to N(μ,1).

Double Exponential(μ, σ) or Laplace

pdf (& support)	$f(x\|\mu,\sigma) = \frac{1}{2\sigma}e^{-\|x-\mu\|/\sigma}$ for $-\infty < x < \infty$
parameter space	$-\infty < \mu < \infty$ (location), $\sigma > 0$ (scale)
mean & variance	$E(X) = \mu$ Var$(X) = 2\sigma^2$
mgf	$M_X(t) = \frac{e^{\mu t}}{1-(\sigma t)^2}$, $\|t\| < \frac{1}{\sigma}$
graph	
shape	A two-sided exponential curve. μ affects the location of the peak, and a higher σ lowers the peak and makes the graph more disperse.
common uses	Could be used to model errors or difference in failure times. In LASSO Bayesian regression, the beta coefficients can be interpreted to have Double Exponential Priors.
R functions	No built-in R functions exist
special cases & relationships	- The difference of two exponential(λ) is Laplace(0, $\frac{1}{\lambda}$) - The absolute value of a Double Exponential(0,σ) is exponential(σ^{-1})
Random Generation	Draw $Y_1, Y_2 \sim$ Exp($\frac{1}{\sigma}$), then $X = Y_1 - Y_2 + \mu$.

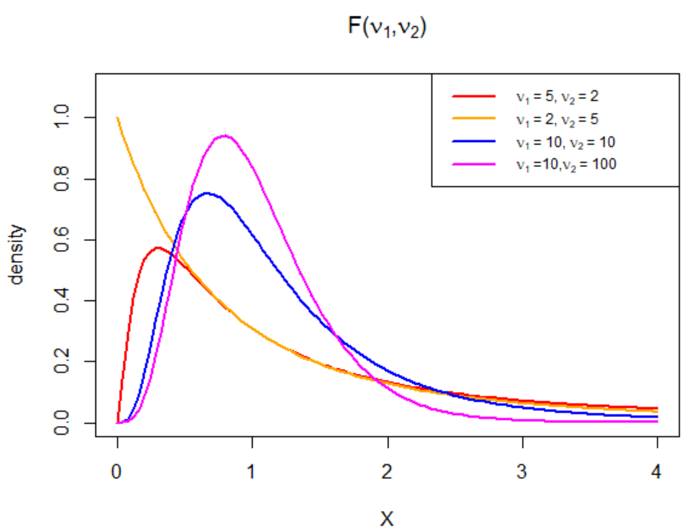
Exponential(λ)

Parameterization	λ **(rate)**	β **(scale)**
pdf (& support)	$f(x\|\lambda) = \lambda e^{-\lambda x}$ for $x > 0$	$f(x\|\beta) = \frac{1}{\beta} e^{\frac{-x}{\beta}}$ for $x > 0$
parameter space	$\lambda > 0$	$\beta > 0$
mean & variance	$E(X) = \frac{1}{\lambda}$ $\quad$ $\text{Var}(X) = \frac{1}{\lambda^2}$	$E(X) = \beta$ $\quad$ $\text{Var}(X) = \beta^2$
mgf	$M_X(t) = \frac{\lambda}{1-\lambda t}, \quad t < \lambda$	$M_X(t) = \frac{1}{1-\beta t}, \quad t < \frac{1}{\beta}$
graph		
shape	(rate) As λ increases, failure rate increases, so density gathers closer to zero. If λ is small, graph is more disperse. Mode is always at zero.	
common uses	Used to model failure times for objects with a constant failure rate. Also, it makes a simple prior for positive parameters.	

R functions	Note: R uses the rate parameterization
	`dexp(x, r)` `pexp(x, r)`
	`qexp(p, r)` `rexp(n, r)`	

special cases & relationships	- Special case of Gamma when $\alpha = 1$ and Weibull when $\gamma = 1$ - Memory-less property: $P(X > t + a\|X > t) = P(X > a)$ - Minimum of Exponential(λ_i) variables is also Exp($\sum_i \lambda_i$) - Sum of n Exp(λ) variabules is Gamma(n, λ)	
Random Generation	Draw $U \sim$ unif(0,1). X = $\frac{-\ln U}{\lambda}$	

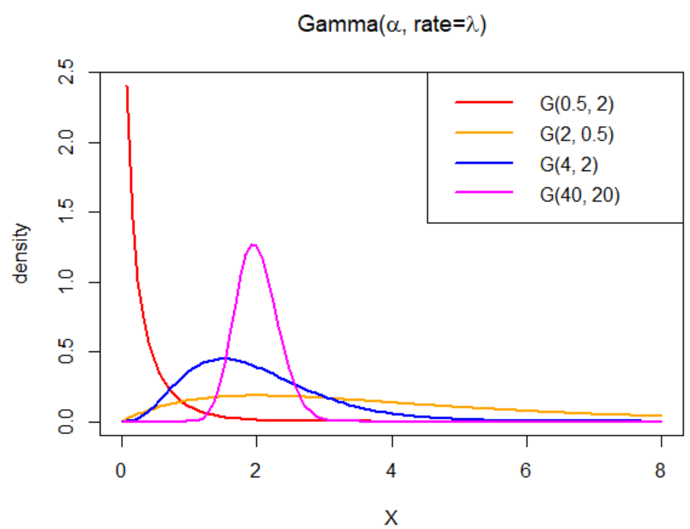
F(ν_1, ν_2)

pdf (& support)	$f(x\|\nu_1,\nu_2) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}(\frac{\nu_1}{\nu_2})^{\nu_1/2}\frac{x^{(\nu_1-2)/2}}{(1+\frac{\nu_1}{\nu_2}x)^{(\nu_1+\nu_2)/2}}$ for $x \geq 0$
parameter space	$\nu_1 > 0; \nu_2 > 0$ (Numerator; Denominator degrees of freedom)
mean & variance	$E(X) = \frac{\nu_2}{\nu_2-2}, \quad \nu_2 > 2$ $\qquad \text{Var}(X) = 2(\frac{\nu_2}{\nu_2-2})^2\frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)}, \quad \nu_2 > 4$
mgf	mgf does not exist, but raw moments can be calculated as: $E(X^n) = \frac{\Gamma(\frac{\nu_1+2n}{2})\Gamma(\frac{\nu_2-2n}{2})}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}(\frac{\nu_2}{\nu_1})^n, \quad n < \frac{\nu_2}{2}$
graph	
shape	If $\nu_1 < 2$ it has an asymptote at 0, $\nu_1 > 2$ the density is shifted away from zero and it is right skewed. As ν_2 increases, graph appears less skewed and if both ν_1, ν_2 are large, it appears normal
common uses	Used for ANOVA tests or likelihood ratio tests and tests with contrasts.

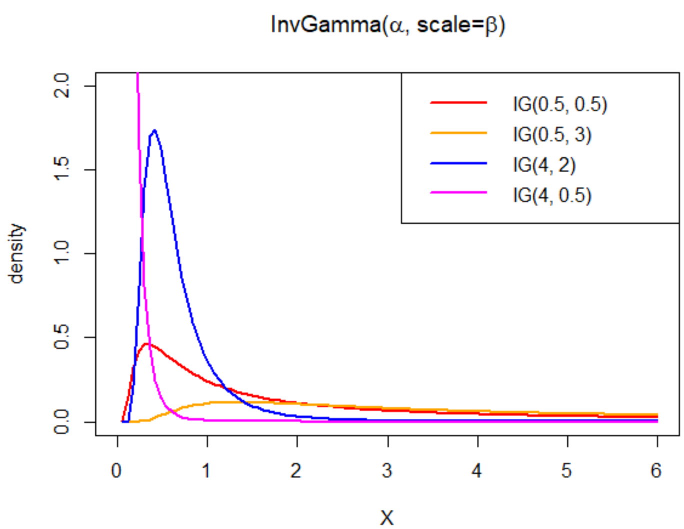
R functions	`df(x,`ν_1`, `ν_2`, ncp)` `pf(x,`ν_1`, `ν_2`, ncp)`
	 `qf(p,`ν_1`, `ν_2`, ncp)` `rf(x,`ν_1`, `ν_2`, ncp)`
	 Where ncp is the non-centrality parameter, which is the ncp for the numerator χ^2

special cases & relationships	- If $U_1 \sim \chi^2_{\nu_1}$ and $U_2 \sim \chi^2_{\nu_2}$, then $(U_1/\nu_1)/U_2/\nu_2) \sim F(\nu_1,\nu_2)$ - If $X \sim F(\nu_1,\nu_2)$ then $1/X \sim F(\nu_2,\nu_1)$ - If $X \sim t_{\nu_2}$, then $X^2 \sim F(1,\nu_2)$ or $\frac{1}{X^2} \sim F(\nu_2, 1)$ - If $X \sim F(\nu_1,\nu_2)$ then $\frac{(\nu_1/\nu_2)X}{1+(\nu_1/\nu_2)X} \sim Beta(\nu_1/2,\nu_2/2)$
Random Generation	Draw $U_1 \sim \chi^2_{\nu_1}$ and $U_2 \sim \chi^2_{\nu_2}$. X $= (U_1/\nu_1)/U_2/\nu_2) \sim F(\nu_1,\nu_2)$

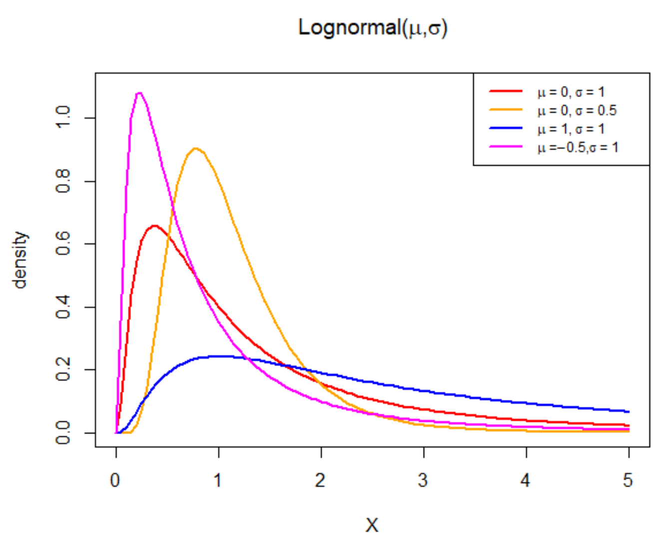
Gamma(α, λ)

<table>
<tr><td>Parameterization</td><td>λ (rate)</td><td>β (scale)</td></tr>
<tr><td>pdf (& support)</td><td>$f(x|\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ for $x > 0$</td><td>$f(x|\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$ for $x > 0$</td></tr>
<tr><td>parameter space</td><td>$\alpha > 0$ (shape); $\lambda > 0$ (rate)</td><td>$\alpha > 0$ (shape); $\beta > 0$ (scale)</td></tr>
<tr><td>mean & variance</td><td>$E(X) = \frac{\alpha}{\lambda}$ $\text{Var}(X) = \frac{\alpha}{\lambda^2}$</td><td>$E(X) = \alpha\beta$ $\text{Var}(X) = \alpha\beta^2$</td></tr>
<tr><td>mgf</td><td>$M_X(t) = (\frac{\lambda}{1-\lambda t})^\alpha, \quad t < \lambda$</td><td>$M_X(t) = (\frac{1}{1-\beta t})^\alpha, \quad t < \frac{1}{\beta}$</td></tr>
<tr><td>graph</td><td colspan="2">
</td></tr>
<tr><td>shape</td><td colspan="2">If $\alpha < 1$, curve has an asymptote at zero, if $\alpha = 1$ it's exponential.
If $\alpha > 1$, mode is pushed away from zero and the curve is right skewed
As λ increases, failure rate increases, so density gathers closer to zero. If λ is small, graph is more disperse.</td></tr>
<tr><td>common uses</td><td colspan="2">Useful for continuous data/priors where the support is positive.</td></tr>
<tr><td>R functions</td><td colspan="2">Note: R uses the <u>rate</u> parameterization
<code>dgamma(x, shape, rate)</code>   <code>pgamma(x, shape, rate)</code>
<code>qgamma(p, shape, rate)</code>   <code>rgamma(n, shape, rate)</code>
Some other useful functions are <code>gamma(x)</code>, <code>lgamma(x)</code>, <code>digamma(x)</code>, and <code>trigamma(x)</code></td></tr>
<tr><td>special cases & relationships</td><td colspan="2">- Reduces to $Exponential(\lambda)$ if $\alpha = 1$; and $\chi^2(k)$ if $\alpha = k/2$ and $\lambda = 1/2$
- As $\alpha \to \infty$ it approaches a Normal distribution
- Sum of n Exp(λ) variabules is Gamma(n, λ)</td></tr>
<tr><td>Random Generation</td><td colspan="2">For integer α, Draw $U_1, ..., U_\alpha \sim$ unif(0,1). X $= \frac{-1}{\lambda} \sum_{i=1}^{\alpha} \ln U_i$ (Sum of α Exponentials)</td></tr>
</table>

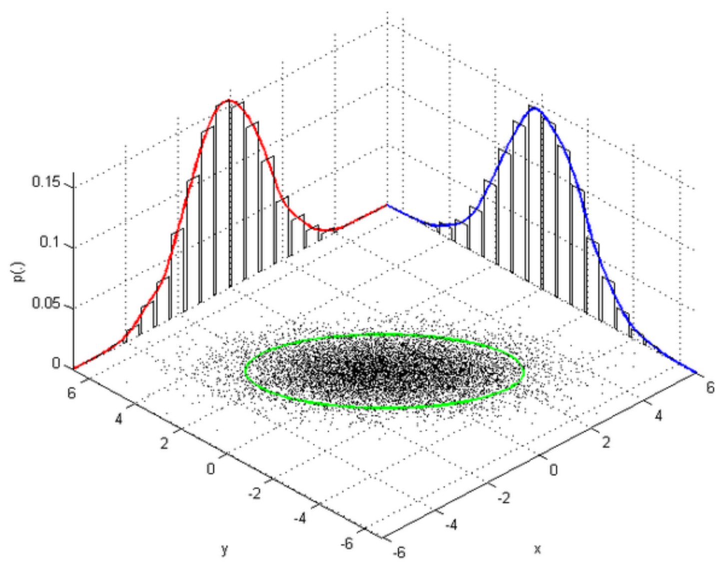
InvGamma(α, λ)

Parameterization	λ **(rate)**	β **(scale)**
pdf (& support)	$f(x\|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)\lambda^\alpha} x^{-\alpha-1} e^{-1/(\lambda x)}$ for $x > 0$	$f(x\|\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$ for $x > 0$
parameter space	$\alpha > 0$ (shape); $\lambda > 0$ (rate)	$\alpha > 0$ (shape); $\beta > 0$ (scale)
mean & variance	$E(X) = \frac{1}{\lambda(\alpha-1)}$ $\quad$ $V(X) = \frac{1}{\lambda^2(\alpha-1)^2(\alpha-2)}$ $\alpha > 1$ for finite mean; $\alpha > 2$ for finite variance	$E(X) = \frac{\beta}{(\alpha-1)}$ $\quad$ $V(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ same restrictions for α apply
mgf	Does not exist	Does not exist
graph		
shape	As α increases, curve gets more dense for lower values of x. As β increases (or λ decreases) density is pushed out and the curve appears more flat.	
common uses	It is conjugate with a normal likelihood. Can model precision (1/variance) well due to positive support.	
R functions	No built-in R functions exist, but the package "invgamma" uses the scale parameterization.	
special cases & relationships	- If Y $Gamma(\alpha, rate = \lambda)$ then $\frac{1}{Y} \sim InvGamma(\alpha, scale = \lambda)$ - CAREFUL: The InvGamma parameterization switches from the Gamma distribution.	
Random Generation	Draw $Y \sim Gamma(\alpha, rate = beta)$. Then $X = 1/Y \sim InvGamma(\alpha, scale = \beta)$	

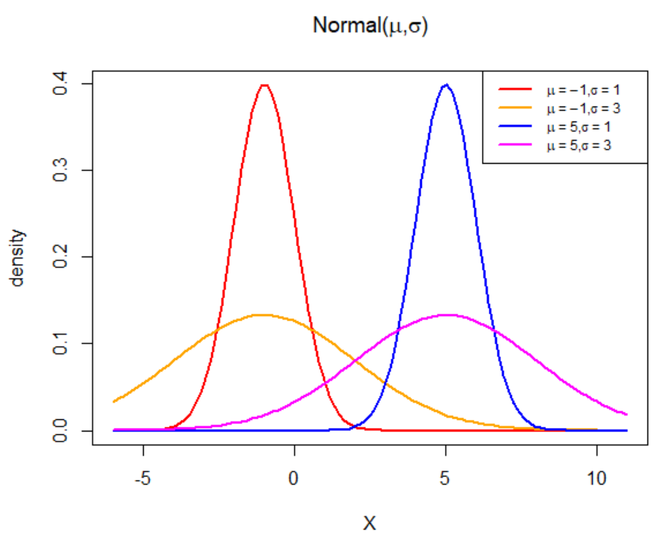
Lognormal(μ, σ^2)

pdf (& support)	$f(x\|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\ln(x)-\mu)^2/(2\sigma^2)}$ for $x \geq 0$
parameter space	$-\infty < \mu < \infty$; $\sigma > 0$; parameters of $\ln(X) \sim Normal(\mu, \sigma^2)$
mean & variance	$E(X) = e^{\mu+\sigma^2/2}$ $\text{Var}(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$
mgf	mgf does not exist, but raw moments can be calculated: $E(X^n) = e^{n\mu+n^2\sigma^2/2}$
graph	
shape	Higher values of σ^2 quickly skew graph right, low values of σ^2 make graph look more normal. Higher values of μ scale/push the curve away from zero.
common uses	Used as some priors with positive support and sometimes errors in linear models become lognormal after transformations. Can be used as an approximation for a product of RVs by the multiplicative version of the central limit theorem (due to the RVs being added on the log scale).
R functions	`dlnorm(x,` μ`,` σ`)` `plnorm(x,` μ`,` σ`)` `qlnorm(p,` μ`,` σ`)` `rlnorm(n,` μ`,` σ`)`
special cases & relationships	- If $X \sim LogNorm(\mu, \sigma^2)$, then $\ln(X) \sim N(\mu, \sigma^2)$ (It's logarithim is normally distributed) - Product of Lognormals is also Lognormal: If $X_i \sim LN(\mu_i, \sigma_i^2)$, then $\prod_{i=1}^{n} X_i \sim LN(\sum \mu_i, \sum \sigma_i^2)$
Random Generation	Draw $Y \sim Normal(\mu, \sigma^2)$. Then, $X = e^Y$

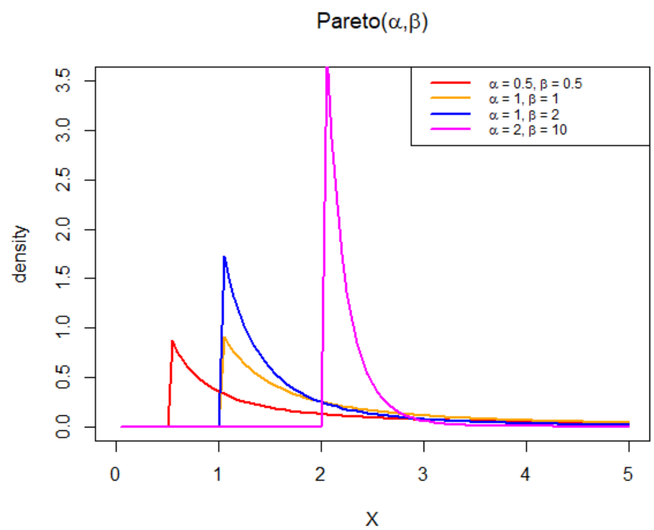
Multivariate Normal($\boldsymbol{\mu}, \mathbf{V}$)

pdf (& support)	$f(\mathbf{x}\|\boldsymbol{\mu}, \mathbf{V}, n) = (2\pi)^{-n/2}\|\mathbf{V}\|^{-1/2} exp(-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{V}^{-1}(\mathbf{Y}-\boldsymbol{\mu}))$ for $x \geq 0$
parameter space	$-\infty < \boldsymbol{\mu} < \infty$; $\mathbf{V}$ is positive semi-definite ($\mathbf{y}^{'}\mathbf{V}\mathbf{y} \geq 0$ for all $\mathbf{y}$) n (dimension) is a positive integer
mean & variance	$E(X) = \boldsymbol{\mu}$ (n x 1 vector) Var$(X) = \mathbf{V}$ (n x n matrix)
mgf	$M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^{\mathbf{T}}\boldsymbol{\mu}+\mathbf{t}^{\mathbf{T}}\mathbf{V}\mathbf{t}/2}$
graph (when $n = 2$)	
shape	Looks like a 2D bell or galaxy cloud. $\boldsymbol{\mu}$ affects location of the cloud. Diagonals of $\mathbf{V}$ affect spread, and off-diagonals (covariances) of $\mathbf{V}$ affect how correlated the dimensions are.
common uses	Simple linear regression is treating one dimension as y and conditioning on all others. Can be used to help describe correlated datasets.
R functions	No built in R functions exist, but package "mvtnorm" can be installed.
special cases & relationships	- When $n = 2$, it's bivariate normal. If $n = 1$, it's a univariate normal - Any linear combination, marginal, joint subset, or conditioning of a MVN is also MVN. - Marginal Variabules are independent iff their covariances are zero.
Random Generation	$X = \boldsymbol{\mu} + \mathbf{C}^{'}\mathbf{z}$, where $\mathbf{z}_{1,..,n} \sim Normal(0, 1)$ and $\mathbf{C}$ is the cholesky decomposition of $\mathbf{V}$

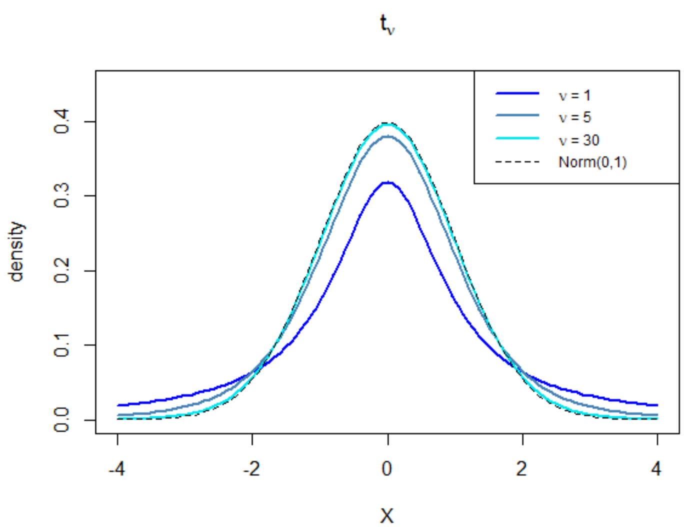
Normal(μ, σ^2)

pdf (& support)	$f(x\|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ for $-\infty < x < \infty$
parameter space	$-\infty < \mu < \infty; \quad \sigma > 0$
mean & variance	$E(X) = \mu$ $\quad$ $\text{Var}(X) = \sigma^2$
mgf	$M_X(t) = e^{t\mu + t^2\sigma^2/2}$
graph	
shape	Bell-shaped. μ affects the centering. Higher values of σ^2 quickly spread the curve out, low values of σ^2 make graph look tighter.
common uses	Many variables and distributions are approximately normal due to the Central Limit theorem Many techniques such as regression, ANOVA, and t-tests assume normality. White noise or errors can often be approximated by a normal distribution. Commonly used as priors for beta coefficients in Bayesian regression.
R functions	`dnorm(x,` μ, σ`)` `pnorm(x,` μ, σ`)` `qnorm(p,` μ, σ`)` `rnorm(n,` μ, σ`)`
special cases & relationships	- Special case of multivariate normal when $n = 1$ - If $X \sim Norm(\mu, \sigma^2)$, then $e^X \sim LogNorm(\mu, \sigma^2)$ - Any linear combination of Normals is normal - If $Z \sim N(0, 1)$, then $Z^2 \sim \chi^2(1)$, if $\mu \neq 0$, it's non-central. If $\sigma^2 \neq 1$, result is gamma. - Central Limit Theorem: If σ^2 is finite, $\overline{X} \to N(\mu, \frac{\sigma^2}{n})$ and $\sum X_i \to N(n\mu, \sigma^2 n)$ as $n \to \infty$
Random Generation	Use the Box-Muller transform. Draw $U_1, U_2 \sim unif(0, 1)$. $X_1 = \mu + \sigma\sqrt{-2\ln U_1}\cos(2\pi U_2)$, $X_1 = \mu + \sigma\sqrt{-2\ln U_1}\sin(2\pi U_2)$

Pareto(α, β)

pdf (& support)	$f(x\|\mu, \sigma^2) = \frac{\beta\alpha^\beta}{x^{\beta+1}}$ for $\alpha < x < \infty$
parameter space	$\alpha > 0$ "minimum"; $\beta > 0$ (shape)
mean & variance	$E(X) = \frac{\beta\alpha}{\beta-1}$, $\beta > 1$ Var$(X) = \frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)}$, $\beta > 2$
mgf	mgf does not exist, but raw momemnts can be calculated: $E(X^n) = \frac{\beta\alpha^n}{\beta-n}$, $\beta > n$
graph	
shape	Looks like a shifted exponential, α is the minimum value, and curve peaks up to it. Higher values of β taper off quicker.
common uses	Modeling distribution of incomes and other econometrics Expressed more generally as the 80-20 principle (ex. 80% wealth, 20% population). Small number of cases produce a large effect.
R functions	No built-in R functions exist, but it has a closed form CDF: $1 - (\frac{\alpha}{x})^\beta$, $x \geq \alpha$
special cases & relationships	- If $Y \sim Exp(\text{rate}=\beta)$, then $\alpha e^Y \sim Pareto(\alpha, \beta)$ OR $Y = \ln(\frac{X}{\alpha})$
Random Generation	One way is to use the Inverse CDF. Draw $U \sim unif(0, 1)$. $X = \alpha(U)^{-1/\beta}$

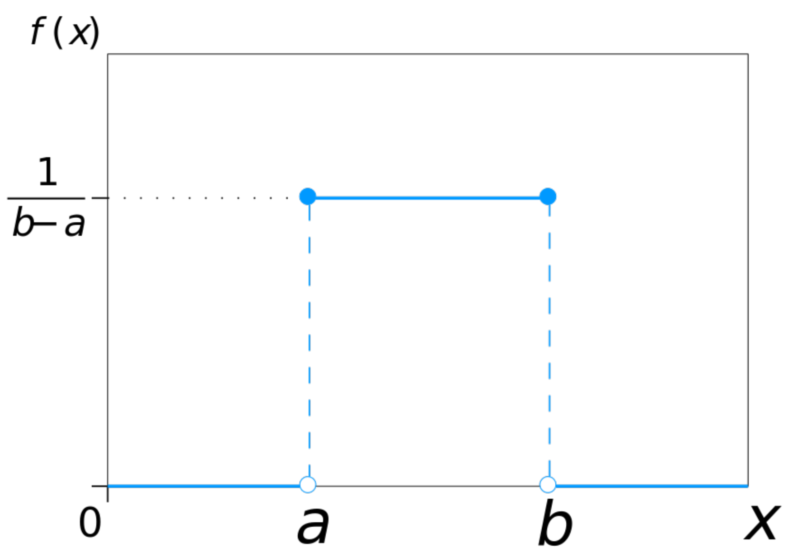
t(ν)

pdf (& support)	$f(x\|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{(1+\frac{x^2}{\nu})^{(\nu+1)/2}}$ for $-\infty < x < \infty$
parameter space	$\nu > 0$ (Degrees of freedom)
mean & variance	$E(X) = 0, \quad \nu > 1$ $\quad$ $\text{Var}(X) = \frac{\nu}{\nu-2}, \quad \nu > 2$
mgf	mgf does not exist, but raw moments can be calculated as: $E(X^n) = \frac{\Gamma(\frac{n+1}{2})\Gamma(\frac{\nu-n}{2})}{\sqrt{\pi}\Gamma(\nu/2)}\nu^{n/2}$, if $n < \nu$ and n is even, OR $E(X^n) = 0$ if $n < \nu$ and n is odd
graph	
shape	Graph has heavier tails than the normal distribution, but as ν gets larger, tails approach the normal's. Always centered at zero and symmetric, unless ncp is non-zero.
common uses	Used for inferences or confidence intervals in a normal population where σ^2 is unknown and s^2 is used instead. Non-Centrality parameter θ is used for power tests where $\theta = \frac{\mu_A - crit^*}{\sigma/\sqrt{n}}$.

R functions	`dt(x, ν, ncp)` `pt(x, ν, ncp)`
	 `qt(p, ν, ncp)` `rt(n, ν, ncp)`
	 Where ncp is the non-centrality parameter, which is the mean of the numerator Normal distribution.

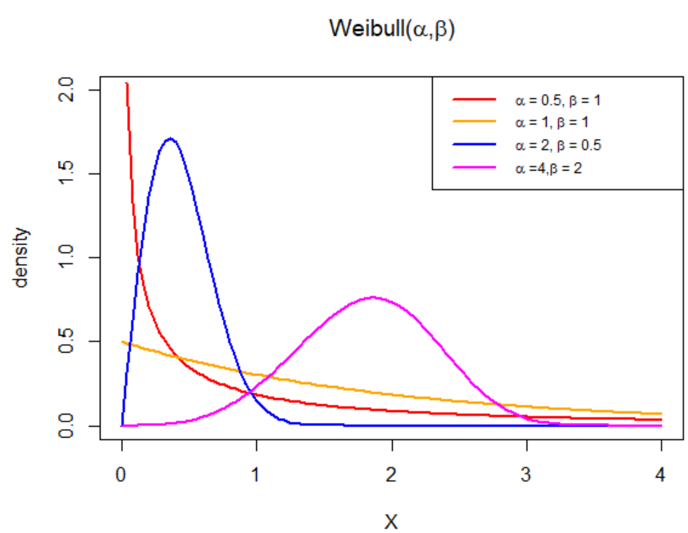
special cases & relationships	- If $U \sim \chi^2_\nu$ and $Z \sim N(0,1)$ are independent, then $\frac{Z}{\sqrt{U/\nu}} \sim t(\nu)$ - If $X \sim t_{\nu_2}$, then $X^2 \sim F(1, \nu_2)$ or $\frac{1}{X^2} \sim F(\nu_2, 1)$ - As $\nu \to \infty$, this approaches a standard normal. - If $X \sim t(\nu)$ then $\frac{\nu}{\nu+X^2} \sim Beta(1/2, \nu/2)$
Random Generation	Draw $U \sim \chi^2_\nu$ and $Z \sim N(0,1)$. Then $X = \frac{Z}{\sqrt{U/\nu}}$.

Uniform(a, b)

pdf (& support)	$f(x\|a,b) = \frac{1}{b-a}$ for $a \leq x \leq b$
parameter space	$-\infty < a < b < \infty$
mean & variance	$E(X) = \frac{b+a}{2}$ $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
mgf	$M_X(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$
graph	
shape	Flat, uniform density on interval $[a, b]$
common uses	Used for uninformative prioirs or modeling continuous outcomes that are all equally likely. Useful for sampling from other distributions by inverse CDF or importance sampling.
R functions	`dunif(x, a, b)` `punif(x, a, b)` `qunif(p, a, b)` `runif(n, a, b)`
special cases & relationships	- Unif$(0, 1)$ is a special case of Beta(α, β) when $\alpha = 1$ and $\beta = 1$ - Sum of two iid uniform variables produces a symmetric triangle distribution
Random Generation	For Unif(a,b), draw $U \sim$ Unif(0,1). $X = a + (b - a)$U

Weibull(α, β)

<table>
<tr><td>Parameterization</td><td>R</td><td>JAGS (From R: $\lambda = (1/\beta)^\alpha$)</td></tr>
<tr><td>pdf (& support)</td><td>$f(x|\alpha, \beta) = \frac{\alpha}{\beta}(\frac{x}{\beta})^{\alpha-1}e^{-(x/\beta)^\alpha}$ for $x \geq 0$</td><td>$f(x|\alpha, \lambda) = \alpha\lambda x^{\alpha-1}e^{-\lambda x^\alpha}$ for $x \geq 0$</td></tr>
<tr><td>parameter space</td><td>$\alpha > 0$ (shape); $\beta > 0$ (scale)</td><td>$\alpha > 0$ (shape); $\lambda > 0$ (rate)</td></tr>
<tr><td>mean &
variance</td><td>$E(X) = \beta\Gamma(1 + 1/\alpha)$
$\text{Var}(X) = \beta^2[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]$</td><td>$E(X) = \lambda^{-1/\alpha}\Gamma(1 + 1/\alpha)$
$\text{Var}(X) = \lambda^{-2/\alpha}[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]$</td></tr>
<tr><td>mgf</td><td>mgf is too complicated, but raw moments are:
$E(X^n) = \beta^n\Gamma(1 + n/\alpha)$</td><td>$E(X^n) = \lambda^{-n/\alpha}\Gamma(1 + n/\alpha)$</td></tr>
<tr><td>graph</td><td colspan="2">



</td></tr>
<tr><td>shape</td><td colspan="2">If $\alpha < 1$, failure rate is decreasing (high infant mortality), so right tail is very long.
If $\alpha = 1$, failure rate is constant, so it's exponential.
If $\alpha > 1$, failure rate increases with time (wear and tear), so graph is left skewed.
As β increases, the shape is the same and the curve spreads out.</td></tr>
<tr><td>common uses</td><td colspan="2">- Used in many reliability, meteorology, engineering, and survival problems due to the shape parameter relating directly to the failure rate.
- Useful for continuous data/priors where the support is positive.</td></tr>
<tr><td>R functions</td><td colspan="2"><code>dweibull(x, alpha, beta)</code>   <code>pweibull(x, alpha, beta))</code>
<code>qweibull(p, alpha, beta)</code>   <code>rweibull(n, alpha, beta)</code></td></tr>
<tr><td>special cases &
relationships</td><td colspan="2">- Reduces to $Exponential(1/\beta)$ if $\alpha = 1$
- If $Y \sim Exp(1/\beta)$, then $\beta(\frac{Y}{\beta})^{1/\alpha} \sim \text{Weibull}(\alpha, \beta)$
- Closed form CDF: $1 - e^{-(x/\beta)^\alpha}$</td></tr>
<tr><td>Random
Generation</td><td colspan="2">Draw $U \sim$ unif(0,1). $X = \beta(-\ln(U))^{1/\alpha}$</td></tr>
</table>