Probability Distributions
Columns (a) through (d)

Distribution	Description	Notes	Notation	Probability Mass or Density Function
				(PMF or PDF)
	(a)	(b)	(c)	(d)
Discrete				
Bernoulli	Takes the value 1 with probability p and takes the value 0 with probability q = 1 - p.	This is a binomial with n = 1 (although this gets circular since the binomial is defined below in terms of the Bernoulli).		$f(x) = p^x q^{1-x}$ for $x \in \{0, 1\}$
Binomial	Sum of n independent Bernoulli trials identically distributed with parameter p.	Converges to a Poisson as n approaches infinity and p approaches zero in such a way that np = λ, the parameter for the Poisson.	B(n, p)	$f(x) = \binom{n}{x} p^x q^{n-x}$ for $x \geq 0$
Geometric	Number of Bernoulli trials with parameter p, until 1 success.	This is equivalent to a negative binomial with r equal to 1.		$f(x) = pq^{x-1}$ for $x \geq 1$
Hypergeometric	Suppose there is an urn with N elements, of which there are A of type 1 and B = N - A of type 2. If n objects are chosen randomly without replacement, the number of elements that are type 1 follows a hypergeometric distribution. Let p = A/N and q = B/N = 1 - p.	A hypergeometric is like a binomial without replacement. When N is very large relative to n, it approaches a binomial.		$f(x) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{N}{n}}$ for $x \geq 0$
Multinomial	An experiment has k possible outcomes with probabilities $p_1, p_2, \ldots, p_k$. If the experiment is repeated n successive times independently, X_i is a random variable of the number of times the experiment results in outcome i.	Expected value and variance can be thought of in terms of a binomial where $p = p_i$ and $q = 1 - p_i$. $\mathrm{Cov}(X_i, X_j) = -np_ip_j$		$f(x_1, x_2, \ldots, x_k)$ $= \binom{n}{x_1, x_2, \ldots, x_k} {p_1}^{x_1} {p_2}^{x_2} \cdots {p_k}^{x_k}$ $= \frac{n!}{x_1!\, x_2! \cdots x_k!} {p_1}^{x_1} {p_2}^{x_2} \cdots {p_k}^{x_k}$ for $\sum_{i=1}^{k} x_i = n$
Negative Binomial	Number of Bernoulli trials with parameter p, until r successes.	Equal to the sum of r geometric distributions.	NB(r, p)	$f(x) = \binom{x-1}{r-1} p^r q^{x-r}$ for $x \geq 1$
Poisson	Expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event.	Can be used as an approximation for the binomial distribution with parameters (n, p) when n is large and p is small so that np is moderate size, with λ = np.	Pois(λ)	$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ for $k \geq 0$

Probability Distributions
Columns (e) through (i)

Distribution	Cumulative Distribution Function	Expected Value	Variance	Median	Moment Generating Function
	(CDF)	(μ)	(σ^2)		(MGF)
	(e)	**(f)**	**(g)**	**(h)**	**(i)**
Discrete					
Bernoulli	$F(x) = \begin{cases} 0 \text{ if } x < 0 \\ q \text{ if } 0 \le x < 1 \\ 1 \text{ if } x \ge 1 \end{cases}$	p	pq		$M(t) = q + pe^t$
Binomial	(expressed in terms of the regularized incomplete beta function)	np	npq		$M(t) = (q + pe^t)^n$
Geometric	$F(x) = 1 - q^x$	$\frac{1}{p}$	$\frac{q}{p^2}$		$M(t) = \frac{pe^t}{1 - qe^t}$
Hypergeometric		np	$npq\left(1 - \frac{n-1}{N-1}\right)$		
Multinomial		$\mathbb{E}[X_i] = np_i$	$\mathrm{Var}(X_i)$ $= np_i(1 - p_i)$ $= np_i q_i$		
Negative Binomial		$\frac{r}{p}$	$\frac{rq}{p^2}$		$M(t) = \left(\frac{pe^t}{1 - qe^t}\right)^r$
Poisson		λ	λ		$M(t) = e^{\lambda(e^t - 1)}$

Probability Distributions
Columns (a) through (d)

Distribution	Description	Notes	Notation	Probability Mass or Density Function
				--------------------(PMF or PDF)--------------------
	(a)	**(b)**	**(c)**	**(d)**
Uniform	A finite number of equally spaced values between a and b are equally likely to be observed; every one of n values has equal probability 1/n, where n equals b - a + 1.		Unif(a, b)	$f(x) = \frac{1}{n}$ for $a \le x \le b$
Continuous				
Chi-Square	A chi-square random variable is parametrized by its degrees of freedom. If Z is a standard normal random variable, then $Z_1^2 + Z_2^2 + \ldots + Z_n^2$ is distributed as a chi-square with n degrees of freedom. It is equivalent to a gamma distribution with α = n / 2 and β = 1 / 2.	The sum of two standard normals, $Z_1^2 + Z_2^2$, is distributed as a chi-square with two degrees of freedom. A chi-square with two degrees of freedom is equal to a gamma with α = 1 and β = 1 / 2. But since α = 1, this is also equal to an exponential with hazard rate λ = 1 / 2 and mean 2.		$f(x) = \frac{1}{2^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}$ for $x > 0$
Exponential	Describes the time between events in a Poisson process with parameter λ. The hazard rate (PDF divided by survival function) of an exponential distribution is constant and equal to λ. $\lambda e^{-\lambda t} / e^{-\lambda t} = \lambda$ Since an exponential distribution's hazard rate is constant and equal to λ, the parameter that defines the distribution, sometimes the distribution is spoken of as being parametrized in terms of its hazard rate.	A Poisson random variable models how many times something will happen in one unit of time (λ is calculated as an average within this time unit). If the occurrences in disjoint time intervals are independent, which is assumed, then the number of occurrences over t units of time is Poisson with mean λt. This can be shown using the fact that the sum of two independent Poisson random variables with parameters λ_1 and λ_2 is distributed as a Poisson with parameter $\lambda = \lambda_1 + \lambda_2$ (a proof of this fact can use the binomial theorem). In the case of t Poissons with the same parameter λ, the resulting random variable is Poisson distributed with parameter λt. The survival function [s(x) = Prob(X > x) = 1 - F(x)] of the exponential distribution equals the probability that the Poisson process with time t is zero, which equals exp(-λt). Note that the probability that the Poisson equals zero is treated as a function of time [i.e., the probability that at time t the Poisson is zero equals exp(-λt)].	Exp(λ)	$f(t) = \lambda e^{-\lambda t}$ for $t > 0$
Gamma 1	The gamma distribution is formed by adding n IID exponential distributions with hazard rate λ.			$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$ for $x > 0$
Gamma 2	For the gamma distribution described, parameter n does not have to be an integer. The more general version of the gamma distribution replaces parameter n with α and λ with β, and α is not constrained to the integers.	$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$		$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$ for $x > 0$

Probability Distributions
Columns (e) through (i)

Distribution	Cumulative Distribution Function	Expected Value	Variance	Median	Moment Generating Function
	------------------(CDF)------------------	-------------(μ)-------------	---------------(σ^2)---------------		----------------(MGF)----------------
	(e)	**(f)**	**(g)**	**(h)**	**(i)**
Uniform		$\frac{1}{2}(a+b)$	$\frac{n^2-1}{12}$	$\frac{1}{2}(a+b)$	$M(t) = \frac{e^t}{n} \cdot \frac{e^{nt}-1}{e^t-1}$
Continuous					
Chi-Square		n	$2n$		$M(t) = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}$
Exponential	$F(t) = 1 - e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\ln 2}{\lambda}$	$M(k) = \frac{\lambda}{\lambda - k}$
Gamma 1		$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$		$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$
Gamma 2		$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$		$M(t) = \left(\frac{\beta}{\beta - t}\right)^\alpha$

Probability Distributions
Columns (a) through (d)

Distribution	Description (a)	Notes (b)	Notation (c)	Probability Mass or Density Function (PMF or PDF) (d)
Lognormal	A probability distribution of a random variable whose natural logarithm is normally distributed. It is parametrized by the mean and variance of its corresponding normal distribution. If X is normally distributed, then exp(X) is lognormally distributed. If Y is lognormally distributed, then ln(Y) is normally distributed.	The lognormal distribution shows up in something almost analogous to the central limit theorem, but for products. Suppose Y is the product of IID random variables: $Y = X_1 X_2 \dots X_n$ Then $\ln(Y) = \ln(X_1) + \ln(X_2) + \dots + \ln(X_n)$ So ln(Y) converges in distribution to a normal as n approaches infinity, from the ordinary central limit theorem. Since ln(Y) converges in distribution to a normal, Y converges in distribution to a lognormal.		$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$ for $x > 0$
Normal	This is the common bell-shaped curve. It arises in a wide variety of places. The sum of n independent, identically distributed (IID) random variables with mean μ and variance σ^2 converges in distribution to a normal distribution with mean $n\mu$ and variance $n\sigma^2$.	Also referred to as a Gaussian distribution.	$N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Pareto 1	Pareto proposed the distribution in terms of its cumulative distribution function, which he believed gave a good approximation to the proportion of incomes that were less than x. The parameter x_m is the minimum value that a Pareto random variable can take, and the parameter α is sometimes referred to as the Pareto index. The lower the value of α, the more uniformly dispersed the distribution (e.g., the larger the Pareto index, the smaller the proportion of very high-income people when the distribution is used to model income). x_m and α must be greater than zero.	If the Pareto index, α, is set to $\log_4 5 = 1.160964$, then the 80-20 rule holds. This rule says that 20% of the population makes 80% of the income. In terms of the distribution, if there are n people, with person i's income equal to X_i, then the value k such that F(k) = 0.8 is the cutoff for which 80% of the population makes less than this income. For the other 20% of the population to make 80% of the income, it must hold that the sum of X_i greater than x, divided by the sum of all X_i is equal to 0.8. I have not seen a proof of why this holds, although I believe it must hold in the limit as n approaches infinity. I have also not seen a derivation of why $\log_4 5$ is the appropriate parameter. I believe this proof and derivation would rely on the fact that the Pareto distribution is a power-law probability distribution, but I'm not sure.		$f(x) = \frac{\alpha x_m^\alpha}{x^{(\alpha+1)}}$ for $x \geq x_m$
Pareto 2	A standard Pareto distribution shifted left so that it starts at 0 instead of x_m (shifted left x units by adding x_m to x in the PDF and CDF equations).	θ was used to parametrize this in place of x_m.		$f(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{(\alpha+1)}}$ for $x \geq 0$
Uniform	Values between a and b are equally likely to be observed; density function is constant across x.		Unif(a, b)	$f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$

Probability Distributions
Columns (e) through (i)

Distribution	Cumulative Distribution Function	Expected Value	Variance	Median	Moment Generating Function
	(CDF)	(μ)	(σ^2)		(MGF)
	(e)	**(f)**	**(g)**	**(h)**	**(i)**
Lognormal		$e^{\mu+\frac{1}{2}\sigma^2}$	$\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$		
Normal		μ	σ^2		$M(t) = e^{\mu t+\frac{\sigma^2 t^2}{2}}$
Pareto 1	$F(x) = 1 - \left(\frac{x_m}{x}\right)^\alpha$ for $x \geq x_m$	$\frac{\alpha x_m}{\alpha - 1}$ for $\alpha > 1$	$\frac{\alpha(x_m^2)}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$		
Pareto 2	$F(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^\alpha$ for $x \geq 0$	$\frac{\theta}{\alpha - 1}$ for $\alpha > 1$	$\frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$		
Uniform		$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	$\frac{1}{2}(a+b)$	$M(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$