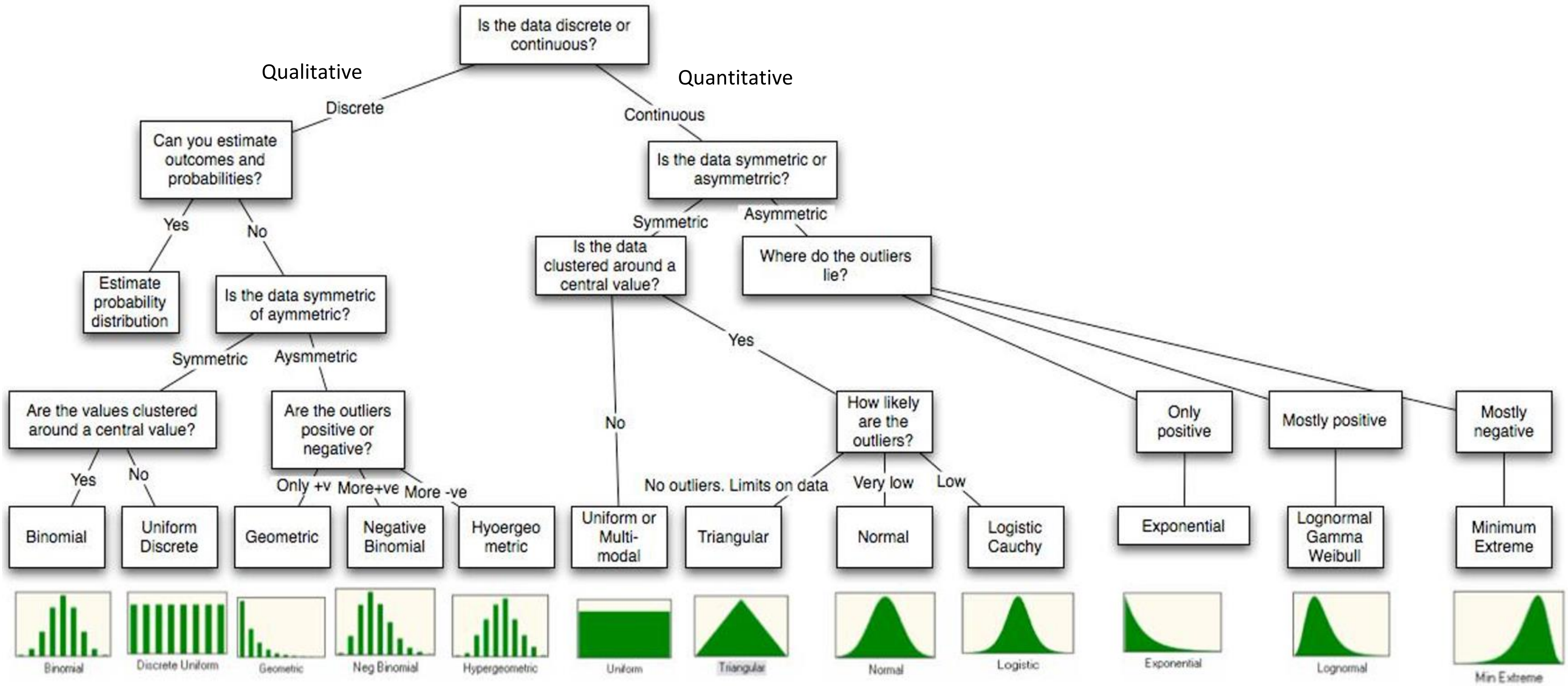
Harold's Statistics
Probability Density Functions
Cheat Sheet

30 May 2016

PDF Selection Tree to Describe a Single Population



Discrete Distributions

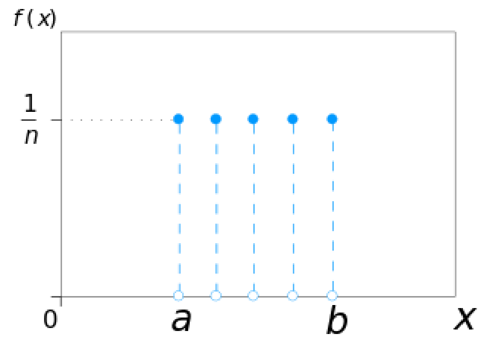
	Notation[1]	$F_X(x)$	$f_X(x)$	$\mathbb{E}[X]$	$\mathbb{V}[X]$	$M_X(s)$
Uniform	$\mathrm{Unif}\{a,\ldots,b\}$	$\begin{cases} 0 & x<a \\ \frac{\lfloor x \rfloor - a+1}{b-a} & a \le x \le b \\ 1 & x>b \end{cases}$	$\frac{I(a<x<b)}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$	$\frac{e^{as}-e^{-(b+1)s}}{s(b-a)}$
Bernoulli	$\mathrm{Bern}(p)$	$(1-p)^{1-x}$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$1-p+pe^s$
Binomial	$\mathrm{Bin}(n,p)$	$I_{1-p}(n-x,x+1)$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$(1-p+pe^s)^n$
Multinomial	$\mathrm{Mult}(n,p)$		$\frac{n!}{x_1!\ldots x_k!}p_1^{x_1}\cdots p_k^{x_k} \quad \sum_{i=1}^{k} x_i = n$	np_i	$np_i(1-p_i)$	$\left(\sum_{i=0}^{k} p_i e^{s_i}\right)^n$
Hypergeometric	$\mathrm{Hyp}(N,m,n)$	$\approx \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$	$\frac{\binom{m}{x}\binom{m-x}{n-x}}{\binom{N}{x}}$	$\frac{nm}{N}$	$\frac{nm(N-n)(N-m)}{N^2(N-1)}$	N/A
Negative Binomial	$\mathrm{NBin}(n,p)$	$I_p(r,x+1)$	$\binom{x+r-1}{r-1}p^r(1-p)^x$	$r\frac{1-p}{p}$	$r\frac{1-p}{p^2}$	$\left(\frac{p}{1-(1-p)e^s}\right)^r$
Geometric	$\mathrm{Geo}(p)$	$1-(1-p)^x \quad x \in \mathbb{N}^+$	$p(1-p)^{x-1} \quad x \in \mathbb{N}^+$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{p}{1-(1-p)e^s}$
Poisson	$\mathrm{Po}(\lambda)$	$e^{-\lambda}\sum_{i=0}^{x}\frac{\lambda^i}{i!}$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ	$e^{\lambda(e^s-1)}$

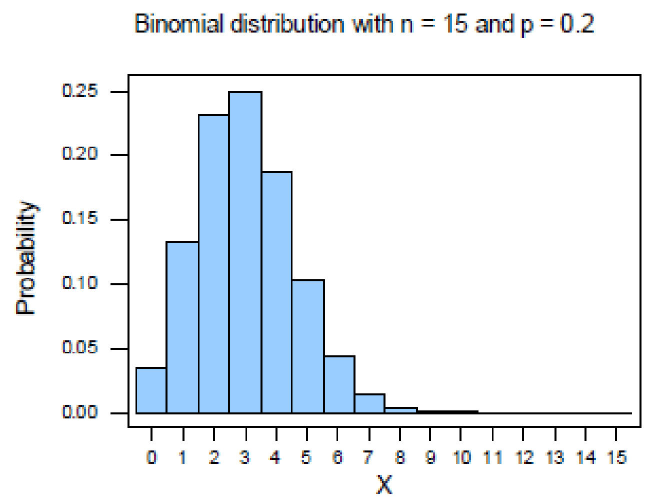
http://www4.ncsu.edu/~swu6/documents/A-probability-and-statistics-cheatsheet.pdf

Continuous Distributions

	Notation	$F_X(x)$	$f_X(x)$	$\mathbb{E}[X]$	$\mathbb{V}[X]$	$M_X(s)$
Uniform	$\mathrm{Unif}(a,b)$	$\begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a<x<b \\ 1 & x>b \end{cases}$	$\frac{I(a<x<b)}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{sb}-e^{sa}}{s(b-a)}$
Normal	$\mathcal{N}(\mu,\sigma^2)$	$\Phi(x)=\int_{-\infty}^{x}\phi(t)\,dt$	$\phi(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	μ	σ^2	$\exp\left\{\mu s+\frac{\sigma^2 s^2}{2}\right\}$
Log-Normal	$\ln\mathcal{N}(\mu,\sigma^2)$	$\frac{1}{2}+\frac{1}{2}\operatorname{erf}\left[\frac{\ln x-\mu}{\sqrt{2\sigma^2}}\right]$	$\frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}$	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	
Multivariate Normal	$\mathrm{MVN}(\mu,\Sigma)$		$(2\pi)^{-k/2}\lvert\Sigma\rvert^{-1/2}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$	μ	Σ	$\exp\left\{\mu^T s+\frac{1}{2}s^T\Sigma s\right\}$
Student's t	$\mathrm{Student}(\nu)$	$I_x\left(\frac{\nu}{2},\frac{\nu}{2}\right)$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}$	0	0	
Chi-square	χ^2_k	$\frac{1}{\Gamma(k/2)}\gamma\left(\frac{k}{2},\frac{x}{2}\right)$	$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2}e^{-x/2}$	k	$2k$	$(1-2s)^{-k/2}\ s<1/2$
F	$\mathrm{F}(d_1,d_2)$	$I_{\frac{d_1x}{d_1x+d_2}}\left(\frac{d_1}{2},\frac{d_1}{2}\right)$	$\frac{\sqrt{\frac{(d_1x)^{d_1}d_2^{d_2}}{(d_1x+d_2)^{d_1+d_2}}}}{x\mathrm{B}\left(\frac{d_1}{2},\frac{d_1}{2}\right)}$	$\frac{d_2}{d_2-2}$	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$	
Exponential	$\mathrm{Exp}(\beta)$	$1-e^{-x/\beta}$	$\frac{1}{\beta}e^{-x/\beta}$	β	β^2	$\frac{1}{1-\beta s}\ (s<1/\beta)$
Gamma	$\mathrm{Gamma}(\alpha,\beta)$	$\frac{\gamma(\alpha,x/\beta)}{\Gamma(\alpha)}$	$\frac{1}{\Gamma(\alpha)\,\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$	$\alpha\beta$	$\alpha\beta^2$	$\left(\frac{1}{1-\beta s}\right)^\alpha\ (s<1/\beta)$
Inverse Gamma	$\mathrm{InvGamma}(\alpha,\beta)$	$\frac{\Gamma\left(\alpha,\frac{\beta}{x}\right)}{\Gamma(\alpha)}$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}e^{-\beta/x}$	$\frac{\beta}{\alpha-1}\ \alpha>1$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)^2}\ \alpha>2$	$\frac{2(-\beta s)^{\alpha/2}}{\Gamma(\alpha)}K_\alpha\left(\sqrt{-4\beta s}\right)$
Dirichlet	$\mathrm{Dir}(\alpha)$		$\frac{\Gamma\left(\sum_{i=1}^k\alpha_i\right)}{\prod_{i=1}^k\Gamma(\alpha_i)}\prod_{i=1}^k x_i^{\alpha_i-1}$	$\frac{\alpha_i}{\sum_{i=1}^k\alpha_i}$	$\frac{\mathbb{E}[X_i](1-\mathbb{E}[X_i])}{\sum_{i=1}^k\alpha_i+1}$	
Beta	$\mathrm{Beta}(\alpha,\beta)$	$I_x(\alpha,\beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{s^k}{k!}$
Weibull	$\mathrm{Weibull}(\lambda,k)$	$1-e^{-(x/\lambda)^k}$	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}e^{-(x/\lambda)^k}$	$\lambda\Gamma\left(1+\frac{1}{k}\right)$	$\lambda^2\Gamma\left(1+\frac{2}{k}\right)-\mu^2$	$\sum_{n=0}^{\infty}\frac{s^n\lambda^n}{n!}\Gamma\left(1+\frac{n}{k}\right)$
Pareto	$\mathrm{Pareto}(x_m,\alpha)$	$1-\left(\frac{x_m}{x}\right)^\alpha\ x\ge x_m$	$\alpha\frac{x_m^\alpha}{x^{\alpha+1}}\ x\ge x_m$	$\frac{\alpha x_m}{\alpha-1}\ \alpha>1$	$\frac{x_m^\alpha}{(\alpha-1)^2(\alpha-2)}\ \alpha>2$	$\alpha(-x_m s)^\alpha\Gamma(-\alpha,-x_m s)\ s<0$

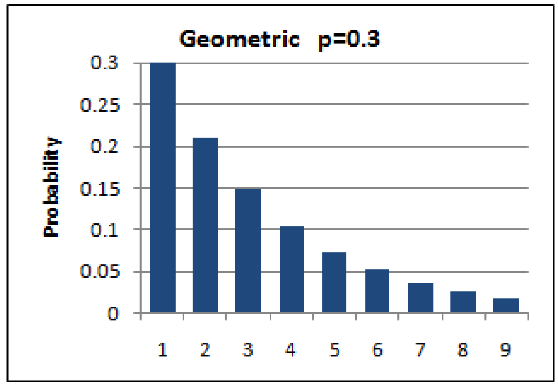
Discrete Probability Density Functions (Qualitative)

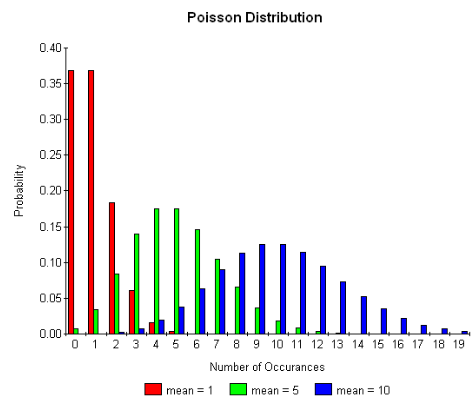
Probability Density Function (PDF)	Mean	Standard Deviation
Uniform Discrete Distribution		
$P(X = x) = \dfrac{1}{b - a + 1}$	$\mu = \dfrac{a + b}{2}$	$\sigma = \sqrt{\dfrac{(b - a)^2}{12}}$
Conditions	• All outcomes are consecutive. • All outcomes are equally likely. • Not common in nature.	
Variables	a = minimum b = maximum	
TI-84	NA	
Example	Tossing a fair die (n = 6)	
Online PDF Calculator	http://www.danielsoper.com/statcalc3/calc.aspx?id=102	

<table>
<tr><th>Probability Density Function (PDF)</th><th>Mean</th><th>Standard Deviation</th></tr>
<tr><td>Binomial Distribution</td><td colspan="2">
</td></tr>
<tr><td rowspan="2">$B(k;n,p) =$
$\boldsymbol{P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}}$
where
$\binom{n}{k} = {}_nC_k = \frac{n!}{(n-k)!\,k!}$</td><td>$\boldsymbol{\mu_x = np}$</td><td>$\boldsymbol{\sigma_x = \sqrt{np(1-p)}} = \sqrt{npq}$
$np \geq 10$ and $nq \geq 10$</td></tr>
<tr><td>$\boldsymbol{\mu_{\hat{p}} = p}$</td><td>$\boldsymbol{\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}} = \sqrt{\frac{pq}{n}}$</td></tr>
<tr><td>$P(X=k) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\frac{(k-np)^2}{npq}}$</td><td colspan="2">Use for large n (> 15) to approximate binomial distribution.</td></tr>
<tr><td>Conditions</td><td colspan="2">• n is fixed.
• The probabilities of success (p) and failure (q) are constant.
• Each trial is independent.</td></tr>
<tr><td>Variables</td><td colspan="2">n = fixed number of trials
p = probability that the designated event occurs on a given trial
(Symmetric if p = 0.5)
X = Total number of times the event occurs $(0 \leq X \leq n)$</td></tr>
<tr><td>TI-84</td><td colspan="2">For one x value:
[2nd] [DISTR] A:binompdf(n,p,x)

For a range of x values [j,k]:
[2nd] [DISTR] A:binompdf([ENTER] n, p, [↓] [↓] [ENTER] [STO>] [2nd] [3] (=L3) [ENTER]
[2nd] [LIST] [→→ MATH] 5:sum(L3,j+1,k+1)</td></tr>
<tr><td>Example</td><td colspan="2">Larry's batting average is 0.260. If he's at bat four times, what is the probability that he gets exactly two hits?

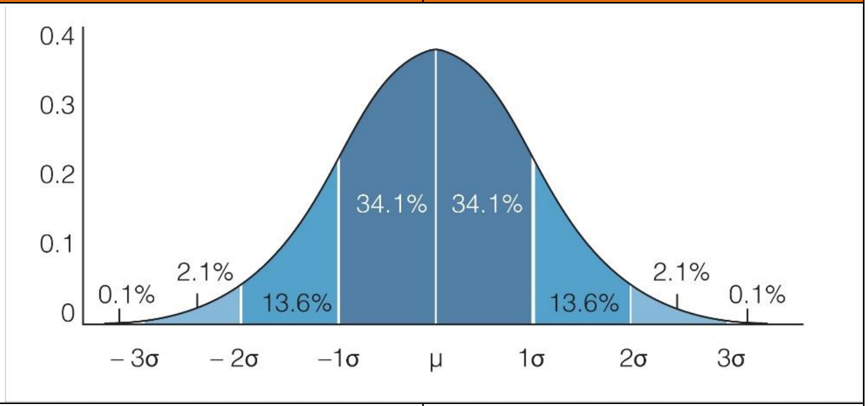
<u>Solution:</u>
n = 4, p = 0.26, x = 2
binompdf(4,0.26,2) = 0.2221 = 22.2%</td></tr>
<tr><td>Online PDF Calculator</td><td colspan="2">http://stattrek.com/online-calculator/binomial.aspx</td></tr>
</table>

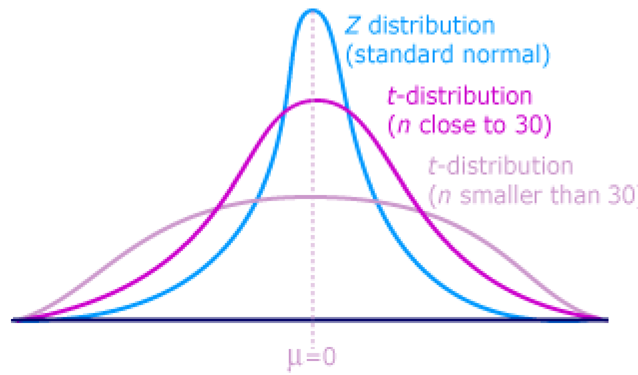
Probability Density Function (PDF)	Mean	Standard Deviation
Geometric Distribution		
$P(X \le x) = q^{x-1}p = (1-p)^{x-1}p$ $P(X > x) = q^x = (1-p)^x$	$\mu = E(X) = \dfrac{\mathbf{1}}{\boldsymbol{p}}$	$\boldsymbol{\sigma} = \dfrac{\sqrt{q}}{p} = \sqrt{\dfrac{\mathbf{1}-\boldsymbol{p}}{\boldsymbol{p}^{\mathbf{2}}}}$
Conditions	• A series of independent trials with the same probability of a given event. • Probability that it takes a specific amount of trials to get a success. Can answer two questions: a) Probability of getting 1st success on the n^{th} trial b) Probability of getting success on $\le n$ trials Since we only count trials until the event occurs the first time, there is no need to count the ${}_nC_x$ arrangements, as in the binomial distribution.	
Variables	*p* = probability that the event occurs on a given trial X = # of trials until the event occurs the 1st time	
TI-84	[2nd] [DISTR] E:geometpdf(p, x) [2nd] [DISTR] F:geometcdf(p, x)	
Example	*Suppose that a car with a bad starter can be started 90% of the time by turning on the ignition. What is the probability that it will take three tries to get the car started?* Solution: *p* = 0.90, *X* = 3 geometpdf(0.9, 3) = 0.009 = 0.9%	
Online PDF Calculator	http://www.calcul.com/show/calculator/geometric-distribution	

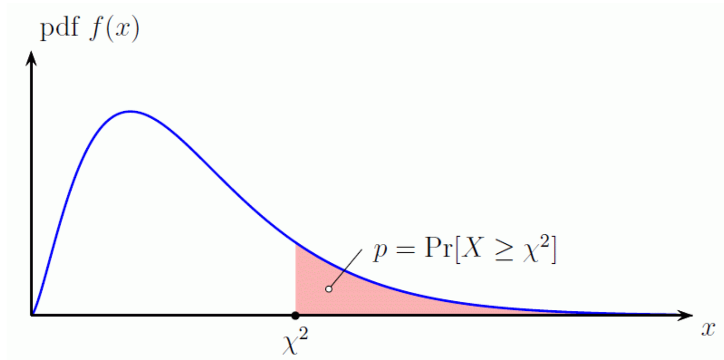
Probability Density Function (PDF)	Mean	Standard Deviation
Poisson Distribution		
$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2,3,4, \ldots$	$\mu = E(X) = \lambda$	$\sigma = \sqrt{\lambda}$
Conditions	• Events occur independently, at some average rate per interval of time/space.	
Variables	λ = average rate X = total number of times the event occurs There is no upper limit on X	
TI-84	[2nd] [DISTR] C:poissonpdf(λ, X) [2nd] [DISTR] D:poissoncdf(λ, X)	
Example	*Suppose that a household receives, on the average, 9.5 telemarketing calls per week. We want to find the probability that the household receives 6 calls this week.* Solution: λ = 9.5, X = 6 poissonpdf(9.5, 6) = 0.0764 = 7.64%	
Online PDF Calculator	http://stattrek.com/online-calculator/poisson.aspx	

Bernoulli	See http://www4.ncsu.edu/~swu6/documents/A-probability-and-statistics-cheatsheet.pdf
tnomial	
Hypergeometric	
Negative Binomial	

Continuous Probability <u>Density</u> Functions (Quantitative)

Probability Density Function (PDF)	**Mean**	**Standard Deviation**
Normal Distribution (Gaussian) / Bell Curve		
$\mathcal{N}(x;\mu,\sigma^2) =$ $\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\mu = \mu$	$\sigma = \sigma$
Special Case: Standard Normal $\mathcal{N}(x;0,1)$	$\mu = 0$	$\sigma = 1$
Conditions	• Symmetric, unbounded, bell-shaped. • No data is perfectly normal. Instead, a distribution is approximately normal.	
Variables	μ = mean σ = standard deviation x = observed value	
TI-84	Have scores, need area: z-scores: [2nd] [DISTR] 1:normalpdf(z, 0, 1) x-scores: [2nd] [DISTR] 1:normalpdf(x, μ, σ) Have boundaries, need area: z-scores: [2nd] [DISTR] 2:normalcdf(left-bound, right-bound) x-scores: [2nd] [DISTR] 2:normalcdf(left-bound, right-bound, μ, σ) Have area, need boundary: z-scores: [2nd] [DISTR] 3:invNorm(area to left) x-scores: [2nd] [DISTR] 3:invNorm(area to left, μ, σ)	
Example	*Suppose the mean score on the math SAT is 500 and the standard deviation is 100. What proportion of test takers earn a score between 650 and 700?* <u>Solution:</u> left-boundary = 650, right boundary =700, μ = 500, σ = 100 normalcdf(650, 700, 500, 100) = 0.0441 = ~4.4%	
Online PDF Calculator	http://davidmlane.com/normal.html	

Probability Density Function (PDF)	Mean	Standard Deviation
Student’s t Distribution	 This distribution was first studied by William Gosset, who published under the pseudonym *Student*.	
Degrees of Freedom	ν = *df* = degrees of freedom = $n-1$ A positive whole number that indicates the lack of restrictions in our calculations. The number of values in a calculation that we can vary. e.g. $\nu = df = 1$ means 1 equation 2 unknowns $\lim_{\nu\to\infty} tpdf(x,\nu) = normalpdf(x)$	
$P(\nu) = \dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\dfrac{x^2}{\nu}\right)^{\frac{-(\nu+1)}{2}}$	$\mu = E(x) = 0$ (always)	$\sigma = 0$
Where the Gamma function $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$	$\Gamma(n) = (n-1)!$	$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
Conditions	• Used for inference about means (Use χ^2 for variance). • Are typically used with small sample sizes or when the population standard deviation isn’t known. • Similar in shape to normal.	
Variables	x = observed value ν = *df* = degrees of freedom = $n-1$	
TI-84	[2nd] [DISTR] 5:tpdf(x, ν) [2nd] [DISTR] 6:tcdf(-∞, t, ν)	
Example	*Suppose scores on an IQ test are normally distributed, with a population mean of 100. Suppose 20 people are randomly selected and tested. The standard deviation in the sample group is 15. What is the probability that the average test score in the sample group will be at most 110?* Solution: n=20, df=20-1=19, μ = 100, $\bar{x}$=110, s = 15 tcdf(-1E99, (110-100)/(15/sqrt(20)), 19) = 0.996 = ~99.6%	
Online PDF Calculator	http://keisan.casio.com/exec/system/1180573204	

Probability Density Function (PDF)	Mean	Standard Deviation
Chi-Square Distribution	 Skewed-right (above) have fewer values to the right, and median < mean.	
$\chi^2(x,k) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{\frac{-x}{2}}$	$\mu = E(X) = k$ $\mu = \sqrt{2}\Gamma\frac{\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$ Mode = $\sqrt{k-1}$	$\sigma^2 = \sqrt{2k}$ $\sigma^2 = k - \frac{2\Gamma\left(\frac{k+1}{2}\right)^2}{\Gamma\left(\frac{k}{2}\right)^2}$
Conditions	• Used for inference about variance in categorical distributions. • Used when we want to test the independence, homogeneity, and "goodness of fit" to a distribution. • Used for counted data.	
Variables	x = observed value ν = *df* = degrees of freedom = $n-1$	
TI-84	[2nd] [DISTR] 7: χ^2pdf(x, ν)	
Example	χ^2pdf() is only used to graph the function.	
Online PDF Calculator	http://calculator.tutorvista.com/chi-square-calculator.html	

Uniform	See http://www4.ncsu.edu/~swu6/documents/A-probability-and-statistics-cheatsheet.pdf
Log-Normal	
Multivariate Normal	
F	
Exponential	
Gamma	
Inverse Gamma	
Dirichlet	
Beta	
Weibull	
Pareto	

Continuous Probability Distribution Functions

Cumulative Distribution Function (CDF)	Mean	Standard Deviation
$P(X \leq x) = \int_{-\infty}^{x} f(x)\, dx$	If $f(x) = \Phi(x)$ (the Normal PDF), then no exact solution is known. Use z-tables or web calculator (http://davidmlane.com/normal.html).	
$\int_{-\infty}^{\infty} f(x)\, dx = 1$	The area under the curve is always equal to exactly 1 (100% probability).	
Integral of PDF = CDF (Distribution)	$F(x) = \int_{-\infty}^{x} f(x)\, dx$	Use the density function $f(x)$, not the distribution function $F(x)$, to calculate $E(X), Var(X)$ and $\sigma(X)$.
Derivative of CDF = PDF (Density)	$f(x) = \frac{dF(x)}{dx}$	
Expected Value	$E(X) = \int_{a}^{b} x\, f(x)\, dx$	
Needed to calculate Variance	$E(X^2) = \int_{a}^{b} x^2\, f(x)\, dx$	
Variance		$Var(X) = E(X^2) - E(X)^2$
Standard Deviation		$\sigma(X) = \sqrt{Var(X)}$