Basic principle of counting	Event 1: m possible outcomes. Event 2: n possible outcomes & indep from event 1	Total: mn possible outcomes - Can be generalized to more evts
Permutations (Order matters)	Num of diff arrangements of n objs	n!
	Permute n objs, of which n_1 are same objs, n_2 are same objs ... and n_r are same objs	$\frac{n!}{n_1!n_2!\dots n_r!}$
	n men sitting in a circle	(n-1)!
Combination (Order not impt)	Select m objs from n objs when order not impt(e.g. AB, AC, AD, BC, BD, CD)	$\binom{n}{m} = \frac{n!}{(n-m)!m!} = \binom{n}{n-m}$
	Combinatorial arg proof. If obj 1 chosen: $\binom{n-1}{r-1}$ ways of selecting r-1 objs from remaining n-1 objs. If obj 1 not chosen: $\binom{n-1}{r}$ ways of selecting r objs from remaining n-1 objs	$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$, 1 ≤ r ≤ n
	Binomial Theorem. $(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$ $\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \dots$	$\sum_{k=0}^{n}(-1)^k\binom{n}{k} = 0$
	Num of subsets of set with n elems	$\sum_{k=0}^{n}\binom{n}{k} = 2^n$ (let x = y = 1)
Multinomial Theorem	Num of divisions of n distinct objs into r **distinct** grps of size n_1, n_2, ...n_r where $n_1 + n_2 + \dots + n_r = n$ (grp matters)	$\binom{n}{n_1}\binom{n-n_1}{n_2}\dots\binom{n-n_1-\dots-n_{r-1}}{n_r} = \binom{n}{n_1,\ n_2,\dots,\ n_r}$ $\frac{n!}{(n-n_1)!n_1!}\frac{(n-n_1)!}{(n-n_1-n_2)!n_2!}\dots\frac{(n-n_1-\dots-n_{r-1})!}{(n-n_1-n_2-\dots-n_r)!n_r!} = \frac{n!}{n_1!n_2!\dots n_r!}$
	Divide n objs into r grps of m each (grp don't matter) (e.g. AB CD, AC BD, AD BC)	$\frac{\binom{n}{m}\binom{n-m}{m}\dots\binom{m}{m}}{r!} = \frac{\frac{n!}{m!m!\dots m!}}{r!}$ (m! r times)
	$(x_1 + x_2 + \dots + x_r)^n = \sum_{(n_1,\dots,n_r):\, n_1+\dots+n_r=n}\binom{n}{n_1,n_2,\dots,n_r}x_1^{n_1}x_2^{n_2}\dots x_r^{n_r}$	
Num of integer solns of eqns	There are $\binom{n-1}{r-1}$ distinct +ve integer-valued vectors (x_1, x_2, ..., x_r) satisfying $x_1 + \dots + x_r = n$	
	There are $\binom{n+r-1}{r-1}$ distinct non-negative integer-valued vectors (x_1, x_2, ..., x_r) satisfying $x_1 + \dots + x_r = n$ OR n identical objs into r distinct grps	

Sample Space	Experiment: outcome not predictable with certainty Sample space: set of all possible outcomes of the experiment	Event: subset of sample space Just draw venn diagram for everything
	E∪F: event containing all outcomes either in E or F or both E & F E∩F OR EF: event containing all outcomes that are both in E & F E^C: event containing all outcomes not in E (complement)	E∪F = {x: x ∈ E or x ∈ F} E∩F = { x: x ∈ E and x ∈ F } OR ∅ E^C = {x: x ∉ E}
	E⊂F: all outcomes in E are in F (subset)	If E⊂F and F⊂E, then E = F
	Commutative Laws: E∪F = F∪E; E∩F = F∩E	Associative Laws: (E ∪ F) ∪ G = E ∪ (F ∪ G); (EF)G = E(FG)
	Distributive Laws: (E ∪ F) ∩ G = EG ∪ FG; EF ∪ G = (E∪G) ∩ (F∪G)	DeMorgan's Laws: $\left(\bigcup_{i=1}^{n} E_i\right)^C = \bigcap_{i=1}^{n} E_i^C$; $\left(\bigcap_{i=1}^{n} E_i\right)^C = \bigcup_{i=1}^{n} E_i^C$
Axioms of Prob	S = sample space. E = event in S Axiom 1: 0 ≤ P(E) ≤ 1	Axiom 2: P(S) = 1 3: For any seq of mutually exclusive events E_1, E_2,... (i.e. E_iE_j = ∅ when i ≠ j), $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$
Simple Propo-sitions	1. P(∅) = 0	2. $P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$ (when sample space is finite)
	3. Strong law of large nums shows $\frac{n(E)}{n}$ converges to P(E) with prob 1 Using axioms, if experiment is repeated many times, by strong law of large numbers, P(E) = proportion which E will occur	
	4. If sample space is finite, Axiom 3 becomes $P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$ for mutually exclusive events E_1, E_2,...	
	5. 3 axioms are basic properties of relative frequencies	6. Use 3 axioms to check whether given fn P(E) is prob fn
	7. $P(E^C)$ = 1 - P(E)	8. If E ⊂ F, then P(E) ≤ P(F)
	9. P(E∪F) = P(E) + P(F) - P(E∩F)	10. P(E∪F∪G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)
	11. Inclusion-exclusion identity	$P(E_1 \cup E_2 \cup \dots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \dots$ $+ (-1)^{r+1}\sum_{i_1<i_2<\dots<i_r} P(E_{i_1}E_{i_2}\dots E_{i_r}) + \dots + (-1)^{n+1}P(E_1E_2\dots E_n)$
	12i. $P(\bigcup_{i=1}^{n} E_i) \le \sum_{i=1}^{n} P(E_i)$	12ii. $P(\bigcup_{i=1}^{n} E_i) \ge \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_iE_j)$
	12iii. $P(\bigcup_{i=1}^{n} E_i) \le \sum_{i=1}^{n} P(E_i) - \sum_{j<i} P(E_iE_j) + \sum_{k<j<i} P(E_iE_jE_k)$	... result can be generalized further on
Sample spaces with equally likely outcomes	If all events in sample space are equally likely to occur, i.e. S = {e_1, e_2, ..., e_n}. P({e_1}) = P({e_2}) = ... = P({e_n}) Then for any event E, take P(E) = $\frac{num\ of\ outcomes\ in\ E}{num\ of\ outcomes\ in\ S} = \frac{num\ of\ outcomes\ in\ E}{n}$. Since P(.) satisfy all 3 axioms, thus P(.) is a probability fn.	
	$e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots$	$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$
Prob as a cts set fn	Incr seq: $E_1 \subset E_2 \subset E_3 \subset \dots \subset E_n \subset E_{n+1}\dots$ Hence $\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$ Decr seq: $E_1 \supset E_2 \supset \dots \supset E_n \supset E_{n+1} \supset\dots$ And $\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$	If {E_n, n ≥ 1} is either incr or decr seq, then $\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$

Conditional Prob and Reduced Sample Space	If P(F) > 0, P(E|F) = $\frac{P(EF)}{P(F)}$	Finding conditional prob using reduced sample space easier
	Multiplication rule: P(AB) = P(A)P(B|A)	$P(E_1E_2\dots E_n) = P(E_1)P(E_2\|E_1)P(E_3\|E_1E_2)\dots P(E_n\|E_1E_2\dots E_{n-1})$
Thrm of Total prob and Bayes' Thrm	Conditioning formula: P(E) = P(E|F)P(F) + P(E|F^C)P(F^C)	$P(F\|E) = \frac{P(EF)}{P(E)} = \frac{P(F)P(E\|F)}{P(E)}$; $P(F^C\|E) = \frac{P(EF^C)}{P(E)} = \frac{P(F^C)P(E\|F^C)}{P(E)}$
	Thrm of total prob: Suppose F_1, F_2, ..., F_n are mutually exclusive events s.t. $\bigcup_{i=1}^{n} F_i$ = S, then $P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(F_i)P(E\|F_i)$	Bayes Thrm: $P(F_j\|E) = \frac{P(EF_j)}{P(E)} = \frac{P(F_j)P(E\|F_j)}{\sum_{i=1}^{n} P(F_i)P(E\|F_i)}$
Indep Events	E and F are independent ⇔ P(EF) = P(E)P(F) ⇔ P(E|F) = P(E)	E and F are independent ⇒ E and F^C are independent
	E, F and G are indep if P(EF) = P(E)P(F), P(EG) = P(E)P(G), P(FG) = P(F)P(G), P(EFG) = P(E)P(F)P(G)	E, F indep and E, G indep ≠ E, FG indep If E, F, G are indep, then E will be indep of any event formed from F and G
P(E|F) is a prob	Conditional prob satisfies 3 axioms	a. 0 ≤ P(E|F) ≤ 1. b. P(S|E) = 1 c. If E_i, i = 1,2,... are mutually exclusive events, then $P(\bigcup_{i=1}^{\infty} E_i\|F) = \sum_{i=1}^{\infty} P(E_i\|F)$.
	Note: Results that are true for unconditional prob also true for conditional prob	Let Q(E) = P(E|F), then Q(E) is a prob. fn on events of S. Then E.g. $Q(E_1 \cup E_2) = Q(E_1) + Q(E_2) - Q(E_1E_2) \Rightarrow P(E_1 \cup E_2\|F) = P(E_1\|F) + P(E_2\|F) - P(E_1E_2\|F)$ E.g. $Q(E_1) = Q(E_1\|E_2)Q(E_2) + Q(E_1\|E_2^C)Q(E_2^C)$. So $P(E_1\|F) = P(E_1\|E_2F)P(E_2\|F) + P(E_1\|E_2^CF)P(E_2^C\|F)$

Pmf & Cdf	For discrete r.v., Prob mass fn (pmf) or prob fn (pf), p(x) = $p_X(x) = \begin{cases} P(X=x), x = x_1, x_2, \dots \\ 0, otherwise \end{cases}$ For r.v., Cumulative dist fn (cdf), F(x) = $F_X(x)$ = P(X ≤ x), x ∈ $\mathbb{R}$	Properties: 1) $P(X = x_i) \geq 0$ for $i \in \mathbb{Z}^+$. 2) $P(X = x_i) = 0$ for $i \notin \mathbb{Z}^+$. 3) $\sum_x P(X=x) = 1$ To check Bin pmf valid, check p + q = 1. $\because \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n$. For a discrete r.v., $F(a) = \sum_{all\ x \leq a} P(X=x)$		
E(x) or μ	$E(X) = \sum_x xP(X=x)$	For a nonnegative integer-valued r.v. Y, $E(Y) = \sum_{i=1}^{\infty} P(Y \geq i) = \sum_{i=0}^{\infty} P(Y > i)$		
E[g(X)]	$E[g(X)] = \sum_x g(x)p(x)$	E(aX + b) = aE(X) + b	n^{th} moment of X, $E(X^n) = \sum_x x^n p(x)$	k^{th} central moment = $E[(X-\mu)^k]$
Variance	$Var(X) = E(X-\mu)^2 = \sum_{x_i}(x_i - \mu)^2 p(x_i) = E(X^2) - [E(X)]^2 = \sigma^2$	$Var(aX + b) = a^2 Var(X)$	σ = standard deviation (SD)	
Formulas	If X ~ Binomial(n, p) and Y ~ Binomial(n-1, p), then $E(X^k) = np * E[(Y+1)^{k-1}]$	If X ~ Binomial(n,p), $P(X = k) = \frac{(n-k+1)p}{k(1-p)} P(X = k-1)$, k = 1,2,..., n		
Random Variables (r.v.)	Bernoulli	X ~ Be (p)	X can only be success (p) or failure (1-p)	$P(X = x) = \begin{cases} p, & x = 1, 'success' \\ 1-p, x = 0, 'failure' \end{cases}$
	Binomial	X ~ Bin(n, p)	k = num of success in n Bernoulli trials	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, k = 0,1,...,n
	Geometric	X ~ Geo(p)	k = num of Bernoulli trials until success is obtained	$P(X = k) = (1-p)^{k-1}p$, k = 1,2,...
	Negative binomial	X ~ NB(r, p) Geo(p) = NB(1, p)	k = num of Bernoulli trials to obtain r successes	$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$, k ≥ r
	Hypergeometric	X ~ H(n, N, m)	N distinct balls: m red, N-m blue. Choose n balls w/o replacement. k = num of red balls chosen	$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$, k = 0,1,...,n
Poisson r.v.	R.v. X is Poisson with parameter λ if for some $\lambda > 0$, $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, k = 0,1,2... X~Poisson(λ). $P(X = i+1) = \frac{e^{-\lambda}\frac{\lambda^{i+1}}{(i+1)!}}{e^{-\lambda}\frac{\lambda^i}{i!}} = \frac{\lambda}{i+1} P(X = i)$, i = 0,1,... $P(X = 0) = e^{-\lambda}$, $P(X = 1) = \lambda e^{-\lambda}$...	Poisson dist: 1. Approximation to binomial dist w large n and small p 2. Num of events occuring at random at certain points in time 1. If X ~ Binomial(n,p), n is large and p is small, then X~Poisson(λ) approximately, where λ = np. Poisson approximation still valid even if trials are not indep, provided their dependence is weak 2. a) Prob only 1 event occuring in interval of length h = λh + o(h) b) Prob ≥ 2 events occuring in interval of length h = o(h) c) For any integers $j_1, j_2, \dots j_n$ and any set of n nonoverlapping intervals, if E_i = event exactly j_i of events occur in the i^{th} intervals, then $E_1, E_2, \dots E_n$ are indep Let N(t) = num of events occuring in time interval [0,t]. If above 3 assumptions true, then N(t) ~ Poisson(λt), where λ is rate of occurrences of events per unit time.		
o(h)	Notation. Little o h: o(h) stands for any fn f(h) s.t. $\lim_{h \to \infty} \frac{f(h)}{h} = 0$. E.g. h^2	Note. o(h) + o(h) = o(h). Since e.g. $h^2 + h^2 = 2h^2 = o(h)$		
Expected Value of Sum of r.v.	For a r.v. X, let X(s) denote value of X when s ∈ S $E(X) = \sum_i x_i P(X = x_i) = \sum_{s \in S} X(s)p(s)$, where s_i = {s: X(s) = x_i}	For r.v. $X_1, X_2, \dots, X_n$, $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$		

Intro	X is a Continuous r.v. if ∃ a nonnegative fn f, defined ∀ x ∈ (-∞, ∞), having property that $P(X \in B) = \int_B f(x)dx$, f(x) = prob density fn (pdf)				
	Properties of pdf	$P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x)dx = 1$	$P(a \leq X \leq b) = \int_a^b f(x)dx$	$P(X = a) = \int_a^a f(x)dx = 0$	dist fn = $F_X(a)$ = P(X < a) = P(X ≤ a) = $\int_{-\infty}^{a} f(x)dx$; $f(x) = \frac{d}{dx}F(x)$, F(x) = cdf
	Interpretation of pdf at x = f(x)	$P(x < X < x + dx) = \int_x^{x+dx} f(y)dy \approx$ f(x)d(x) (area of rectangle). $f(x) \approx \frac{P(x < X < x+dx)}{dx}$			
	If X ~ cdf F(x), then F(X) = U ~ uniform(0,1)	$X = F^{-1}(U)$ ~ cdf F(x)			
Expectation and Variance	$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$	Lemma. If Y ≥ 0, $E(Y) = \int_0^{\infty} P(Y > y)dy$	If X ~ pdf f(x), $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$		
	Corollary: E(aX + b) = aE(X) + b	$Var(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$			
Uniform r.v.	X ~ Uniform (α, β)	pdf $f(x) = \begin{cases} \frac{1}{\beta - \alpha}, \alpha < x < \beta \\ 0, otherwise \end{cases}$	$E(X) = \frac{\alpha+\beta}{2}$, $Var(X) = \frac{(\beta-\alpha)^2}{12}$	If X ~ Uniform (α, β), then $\frac{X-\alpha}{\beta-\alpha}$ ~ Uniform(0,1)	
Normal r.v.	X ~ N (μ, σ^2)	pdf $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, -∞ < x < ∞	$E(X) = \mu$, $Var(X) = \sigma^2$	If X ~ N (μ, σ^2), then $\frac{X-\mu}{\sigma}$ ~ N(0,1)	
	Standard normal dist: Z ~ N (0,1)	$F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}y^2} dz$	If Y ~ N(μ, σ^2) and a is constant, $F_Y(a) = P(Y \leq a) = P(\frac{Y-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}) = F(\frac{a-\mu}{\sigma})$		
	P(Z ≥ 0) = P(Z ≤ 0) = .5	-Z ~ N(0, 1)	P(Z ≤ x) = 1 - P(Z > x)	P(Z ≤ -x) = P(Z ≥ x)	aZ + b ~ N(b, a^2)
	The normal approximation to the binomial distribution	If S_n ~ Binomial(n,p), then $\frac{S_n - np}{\sqrt{np(1-p)}}$ ~ N(0,1) approximately for large n			
Exponential r.v.	X ~ Exponential(λ) or Exp(λ) for some $\lambda > 0$	pdf $f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, otherwise \end{cases}$	dist fn, $F_X(x) = 1 - e^{-\lambda x}$, x ≥ 0	$P(X > s) = e^{-\lambda s}$	
	Memoryless Property: P(X > s + t | X > t) = P(X > s) ∀ s, t > 0 – (1)	P(X > s + t) = P(X > s)P(X > t) ∀ s, t > 0 (derived from (1))			
Other cts dist	X ~ Gamma(α, λ) where $\alpha > 0$ and $\lambda > 0$ if pdf $f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, x \geq 0 \\ 0, \quad x < 0 \end{cases}$ where gamma fn $\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1} dy$				
	$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, $\alpha > 1$	$\Gamma(1/2) = \sqrt{\pi}$	If events are occurring randomly in time and follows 3 assumptions of poisson r.v. N(t) ~ Poisson(λt), then amt of time one has to wait until total of n events has occurred is a gamma r.v. with parameters (n, λ)		
	If α is an int, say α = n, then $\Gamma(n) = (n-1)!$				
	If X ~ Gamma(1, λ), then X ~ Exponential(λ)				
	If X ~ Gamma(α, λ), then $E(X) = \alpha/\lambda$ and $Var(X) = \alpha/\lambda^2$		X_i ~ Exp(λ), i = 1,2,... and X_i are indep. Then $X_1 + X_2 + \dots + X_n$ ~ Gamma(n, λ). Similar to negative binomial for discrete case		
	If X ~ Gamma($\frac{n}{2}, \frac{1}{2}$), then X ~ Chi-square dist with n deg of freedom				
	X ~ Beta(a,b) if pdf $f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, 0 < x < 1 \\ 0, \quad otherwise \end{cases}$ where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$	$E(X) = \frac{a}{a+b}$, $Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$			
	Beta(1,1) = Uniform(0,1)	$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$	X ~ Cauchy(θ) with $-\infty < \theta < \infty$ if pdf $f(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$, -∞ < x < ∞	$E(X^n)$ DNE for n = 1,2,...	
	Weibull Dist W(ν, a, b), $f_X(x) = \begin{cases} \frac{b}{a}\left(\frac{x-\nu}{a}\right)^{b-1} e^{-\left(\frac{x-\nu}{a}\right)^b}, x > \nu \\ 0, if\ x \leq \nu \end{cases}$	$E(X) = a\Gamma(1+\frac{1}{b})$. $Var(X) = a^2\left[\Gamma\left(1+\frac{2}{b}\right) - \left(\Gamma\left(1+\frac{1}{b}\right)\right)^2\right]$. W(1, λ, 0) = Exp(λ)			
Approximation of Binomial r.v.	Let X ~ Binomial(n,p). Assume n is large(≥30). 1. Normal approximation. Binomial(n,p) ≈ N(np, npq) OR $\frac{X-np}{\sqrt{npq}} \approx Z$, where Z ~ N(0,1). Approximation good if npq ≥ 10 Continuity correction. When finding prob of X using the normal dist P(X = k) = P(k-1/2 < X < k+1/2). P(X ≥ k) = P(X ≥ k-1/2). P(X ≤ k) = P(X ≤ k+1/2)	2. Poisson dist. Used when n is large, p is small and np is moderate. Rule of thumb: use Poisson approximation if p < 0.1 and put λ = np. If p > 0.9, put λ = n(1-p) and work in terms of 'failure'			

Joint Dist Fn	$F(x,y) = P(X \le x, Y \le y)$		Note $\{X > a, Y > b\} \ne \{X \le a, Y \le b\}^C$	
	Marginal cdf of X, $F_X(x) = \lim_{y\to\infty} F(x,y)$		Marginal cdf of Y, $F_Y(y) = \lim_{x\to\infty} F(x,y)$	
	$P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F(a,b)$		$P(a_1 \le X \le a_2, b_1 \le Y \le b_2) = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1)$	
	If X and Y are discrete r.v. then their joint pmf $p(i, j) = P(X = i, Y = j)$			
	Marginal pmf of X, $P(X = i) = \sum_j P(X = i, Y = j)$		Marginal pmf of Y, $P(Y = j) = \sum_i P(X = i, Y = j)$	
	If X and Y are cts r.v., then their joint pdf $P[(X, Y) \in C] = \iint_C f(x,y)\,dxdy$ = vol under the surface f(x, y) over the region C		If $C = \{(x,y): x \in A, y \in B\}$, then $P(X \in A, Y \in B) = \int_B\int_A f(x,y)\,dxdy$	
	Joint cdf $F(a, b) = P(X \in (-\infty, a], Y \in (-\infty, b]) = \int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)\,dxdy$		$f(a, b) = \frac{\partial^2}{\partial a \partial b}F(a, b)$	
	Interpretation of joint pdf f(a, b) (density)		$P(a < X < a + da, b < Y < b + db) = \int_b^{b+db}\int_a^{a+da} f(x,y)\,dxdy \approx f(a,b)\,da\,db$	
	Marginal pdf of X, $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$		Marginal pdf of Y, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$	
Indep r.v.	X and Y are indep: $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$		X and Y indep: $P(X \le a, Y \le b) = P(X \le a)P(Y \le b)$ OR $F(a, b) = F_X(a)F_Y(b)$	
	Discrete case: X and Y indep: $P(X = x, Y = y) = P(X = x)P(Y = y)\ \forall\, x,y$		Cts case: X and Y indep: $f(x, y) = f_X(x)f_Y(y)\ \forall\, x,y$	
	X and Y are indep if knowing value of one does not change dist of other		X and Y indep iff their joint pdf/pmf can be expressed as f(x, y) = g(x)h(y), **$-\infty < x < \infty$, $-\infty < y < \infty$**	
	Independence is a symmetric relation. If X is indep of Y, then Y is indep of X			
Sum of indep r.v.	Suppose X and Y are indep cts r.v., then $F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)\,dy$, $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)\,dy$			
	Dist of sums of indep r.v.	1. $X_i \sim \text{Gamma}(t_i, \lambda)$, i = 1,…,n $\Rightarrow \sum_{i=1}^{n} X_i \sim \text{Gamma}(\sum_{i=1}^{n} t_i, \lambda)$	3. $Z_i \sim N(0, 1)$, i = 1,…,n $\Rightarrow \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$ (chi-square w n deg of freedom)	
		2. $X_i \sim \text{Exp}(\lambda)$, i = 1,…,n $\Rightarrow \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda)$	5. $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2) \Rightarrow X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$	
		4. $X_i \sim N(\mu_i, \sigma_i^2)$, i = 1,…,n $\Rightarrow \sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$	6. $X \sim \text{Binomial}(n, p)$, $Y \sim \text{Binomial}(m, p) \Rightarrow X + Y \sim \text{Binomial}(n + m, p)$	
Conditional dist for Discrete Case	If P(F) > 0, $P(E \mid F) = \frac{P(EF)}{P(F)}$		conditional pmf of X given Y = y is $p_{X\mid Y}(x\mid y) = P(X = x \mid Y = y) = \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{p(x,y)}{p_Y(y)}$, $\forall$ y s.t. $p_Y(y) > 0$, and p(x,y) is joint pmf of X and Y	
	If X is indep of Y, then $p_{X\mid Y}(x\mid y) = p_x(x)$, i.e. $P(X=x \mid Y=y) = P(X=x)\ \forall\, x,y$		Conditional dist fn of X given Y = y is $F_{X\mid Y}(x,y) = P(X \le x \mid Y = y) = \sum_{a\le x} P(X = a \mid Y = y) = \sum_{a \le x} p_{X\mid Y}(a\mid y)$	
Conditional dist for Cts case	If X and Y have joint pdf f(x,y), then conditional pdf of X given Y = y is $f_{X\mid Y}(x\mid y) = \frac{f(x,y)}{f_Y(y)}$, $\forall$ y s.t. $f_Y(y) > 0$			
	Conditional prob/cumulative dist fn of X given Y = y is $F_{X\mid Y}(a\mid y) = P(X \le a \mid Y=y) = \int_{-\infty}^{a} f_{X\mid Y}(x\mid y)dx$		X and Y indep: $f_{X\mid Y}(x\mid y) = f_X(x)$	
Joint Prob Dist of fns of r.v.	Let X_1 and X_2 be jointly cts r.v. w joint pdf $f_{X_1,X_2}(x_1, x_2)$. Suppose $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ and assume Y_1 and Y_2 satisfy: 1. x_1 and x_2 can be uniquely expressed in terms of y_1 and y_2 2. y_1 and y_2 have cts partial derivatives at all point (x_1, x_2) and $J(x_1, x_2) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} = \frac{\partial y_1}{\partial x_1}\frac{\partial y_2}{\partial x_2} - \frac{\partial y_1}{\partial x_2}\frac{\partial y_2}{\partial x_1} \ne 0$ at all point (x_1, x_2) Then Y_1 and Y_2 are jointly cts w joint pdf $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)\frac{1}{\lvert J(x_1,x_2)\rvert}$, where x_1, x_2 are expressed in terms of y_1 y_2			

Expectation of Sums of r.v.	$P(a \le X \le b) \Rightarrow a \le E(X) \le b$		If X and Y have a joint pmf p(x,y), then $E[g(X, Y)] = \sum_y\sum_x g(x,y)p(x,y)$	
	$X \ge Y \Rightarrow E(X) \ge E(Y)$		If X and Y have a joint pdf f(x,y), then $E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy$	
	If E(X) & E(Y) are finite, E(X + Y) = E(X) + E(Y)		If $E(X_i)$ is finite for i = 1,…,n, then $E(X_1 + … + X_n) = E(X_1) + … E(X_n)$	
Covariance, Variance of Sums, Correlations	X and Y indep $\Rightarrow$ E[g(X)h(Y)] = E[g(X)]E[h(Y)]		Cov(X, Y) = E[(X - E[X])(Y - E[Y])] (measure dir[n] of linear relationship btw X and Y)	
	Cov(X, Y) = E(XY) - E(X)E(Y)		X and Y indep $\Rightarrow$ Cov(X, Y) = 0. (opp not necessary; can have non linear r/s)	
	Cov(X, Y) = Cov(Y , X)	Cov(X, X) = Var(X)	Cov(aX, Y) = aCov(X, Y)	$Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$
	$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X_i, Y_j)$		If $X_1,…,X_n$ are pairwise indep, i.e. X_i, X_j indp for i≠j, then $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$	
	Correlation of 2 r.v X and Y, $\rho(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ if Var(X)Var(Y) > 0		ρ measure strength and dir[n] of linear r/s	$-1 \le \rho(X, Y) \le 1$
	$\rho(X, Y) = 1 \Rightarrow Y = a + bX$, $b = \frac{\sigma_Y}{\sigma_X} > 0$	$\rho(X, Y) = -1 \Rightarrow Y = a + bX$, $b = -\frac{\sigma_Y}{\sigma_X} < 0$	X, Y indep $\Rightarrow \rho(X, Y) = 0$. (converse not true)	
	$E(\bar{X}) = \mu$	$Var(\bar{X}) = \frac{\sigma^2}{n}$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$E(s^2) = \sigma^2$
Conditional Expectation	$E(X\mid Y = y) = \sum_x xP(X = x\mid Y = y) = \sum_x xp_{X\mid Y}(x\mid y)$, for $p_Y(y) > 0$		$E(X\mid Y = y) = \int_{-\infty}^{\infty} xf_{X\mid Y}(x\mid y)\,dx$, for $f_Y(y) > 0$	
	$E(g(X)\mid Y = y) = \sum_x g(x)P(X = x\mid Y = y) = \sum_x g(x)p_{X\mid Y}(x\mid y)$		$E(g(X)\mid Y = y) = \int_{-\infty}^{\infty} g(x)f_{X\mid Y}(x\mid y)\,dx$	
	$E(X) = E(E(X\mid Y))$ (wrt Y, wrt $X\mid Y=y$)	$E(X) = \sum_y E(X\mid Y = y)P(Y = y)$	$E(X) = \int_{-\infty}^{\infty} E(X\mid Y = y)f_Y(y)\,dy$	
	$P(A) = \sum_y P(A\mid Y = y)P(Y = y)$. If $F_i = \{Y = y_i\}$. Then $P(A) = \sum_{i=1}^{n} P(A\mid F_i)P(F_i)$		$P(A) = \int_{-\infty}^{\infty} P(A\mid Y = y)f_Y(y)\,dy$	
	Conditional Var, $Var(X\mid Y) = E[(X - E(X\mid Y))^2\mid Y]$	$Var(X\mid Y) = E(X^2\mid Y) - [E(X\mid Y)]^2$	$Var(X) = E[Var(X\mid Y)] + Var[E(X\mid Y)]$	
Moment Generating Functions	Moment Generating Fn: $M(t) = E(e^{tX})$	X is discrete w pmf p(x), $M(t) = \sum_x e^{tx}p(x)$	Y is cts w pdf f(x): $M(t) = \int_{-\infty}^{\infty} e^{tx}f(x)\,dx$	
	$M^n(t) = E(X^n e^{tX})$, n ≥ 1	$M^n(0) = E(X^n)$, n ≥ 1	mgf unique to each distribution, same as pdf and cdf	
	X and Y indep $\Rightarrow M_{X+Y}(t) = M_x(t)M_Y(t)$			

Markov's Inequality	If X is a r.v. that takes only nonnegative values, then for any a > 0, $P(X \ge a) \le \frac{E(X)}{a}$	
Chebyshev's Inequality	If X is a r.v. w finite mean μ and var σ^2, then for any value of k > 0, $P(\lvert X- \mu\rvert \ge k) \le \frac{\sigma^2}{k^2}$	If Var(X) = 0, then P(X = E[X]) = 1
Weak Law of large numbers	Let $X_1, X_2,…$ be a seq of indep and identically distributed r.v. each having finite mean $E[X_i] = \mu$. Then for any $\varepsilon > 0$, $P\left\{\left\lvert\frac{X_1+\cdots+X_n}{n} - \mu\right\rvert \ge \varepsilon\right\} \to 0$ as $n \to \infty$	
Central Limit theorem	Let $X_1, X_2,…$ be a seq of indep and identically distributed r.v. each having mean μ and var σ^2. Then $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \to N(0,1)$ as $n \to \infty$. i.e. for $-\infty < a < \infty$, $P(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \le a) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-\frac{1}{2}y^2}\,dy$ as $n \to \infty$. Note $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} = \frac{\frac{X_1+\cdots+X_n}{n}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}-E(\bar{X})}{\sqrt{Var(\bar{X})}}$ Continuity correction. When using normal dist to approx P(X = k) = P(k-1/2 < X < k+1/2). P(X ≥ k) = P(X ≥ k-1/2). P(X > k) = P(X ≥ k + 1/2). P(X ≤ k) = P(X ≤ k+1/2) P(X < k) = P(X ≤ k - 1/2)	

	Lemma. Let $Z_1, Z_2, \ldots$ be a seq of r.v. having distribution fns F_{Z_n} mgf M_{Z_n}, n ≥ 1 and let Z be a r.v. w dist fn F_Z and mgf M_Z. If $M_{Z_n}(t) \to M_Z(t)$ for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which $F_Z(t)$ is cts
Strong Law of large nums	Let $X_1, X_2, \ldots$ be a seq of indep and identically distributed r.v. each having finite mean $E[X_i] = \mu$. Then w prob 1, $\frac{X_1+\cdots+X_n}{n} \to \mu$ as $n \to \infty$
One-sided Chebyshev's Inequality	X is r.v. w mean 0, var σ^2, $P(X \geq a) \leq \frac{\sigma^2}{\sigma^2+a^2}$
Chernoff bounds	$P(X \geq a) \leq e^{-ta}M(t)$ for all t > 0. $P(X \leq a) \leq e^{-ta}M(t)$ for all t < 0
Jensen's Inequality	A twice-differentiable real-valued fn f(x) is convex if f''(x) ≥ 0 for all x; concave = f''(x) , 0 If f(x) is a convex fn, then E[f(X)] ≥ f(E(X)), if E(X) exists and is finite

	Pmf p(x), pdf f(x)	Mgf M(t)	Mean	Variance
Bernoulli			p	p(1-p)
Binomial w param n,p; 0 ≤ p ≤ 1	$\binom{n}{x}p^x(1-p)^{n-x}$, x = 0,1,…,n	$(pe^t + 1 - p)^n$	np	np(1-p)
Poisson w param $\lambda > 0$	$e^{-\lambda}\frac{\lambda^x}{x!}$, x = 0,1,2…	$e^{\lambda(e^t-1)}$	λ	λ
Geometric w param p; 0 ≤ p ≤ 1	$p(1-p)^{x-1}$, x = 1,2,…	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial w param r, p; 0 ≤ p ≤ 1	$\frac{(x-1)!}{(x-r)!(r-1)!}p^r(1-p)^{x-r}$, n = r, r+1, …	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric	$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$, k = 0,1,…,n		$\frac{nm}{N}$	$\frac{nm(N-m)(N-n)}{N^2(N-1)}$
Discrete uniform (a,b)			$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha+1)^2-1}{12}$
Uniform over (a, b)	$\begin{cases}\frac{1}{b-a}, a<x<b \\ 0, otherwise\end{cases}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential w param $\lambda > 0$	$\begin{cases}\lambda e^{-\lambda x}, x \geq 0 \\ 0, otherwise\end{cases}$	$\frac{\lambda}{\lambda-t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma w param (s, λ), $\lambda > 0$	$\begin{cases}\frac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)}, x \geq 0 \\ 0, \quad x<0\end{cases}$	$\left(\frac{\lambda}{\lambda-t}\right)^s$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$
Normal w param (μ, σ^2)	$\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, -∞ < x < ∞	$e^{t\mu+\frac{\sigma^2t^2}{2}}$	μ	σ^2