Fundamentals:

$x \in S : A \cup B = \{x \in A \text{ or } x \in B\}$
$A \cup B = B \cup A$, $A \cup A = A$
$A \cup \emptyset = A$, $A \cup S = S$
$A \subset B \Rightarrow A \cup B = B$
$A_1, A_2, \ldots, A_n \Rightarrow A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{i=n} A_i$
$\bigcup_{i=1}^{\infty} A_i \to \bigcup_{i \in I} A_i$
$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$
$A \cap B = \{x \in A \text{ and } x \in B\} = AB$
$A \cap B = B \cap A$, $A \cap A = A$
$A \cap \emptyset = \emptyset$, $A \cap S = A$
$A \subset B \Rightarrow A \cap B$
$\bigcap_{i \in I} A_i = \bigcap_{i=1}^{\infty} A_i$
$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$
$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$
$A^c = \{x \in S : x \notin A\}$
$(A^c)^c = A$, $\emptyset^c = S$, $S^c = \emptyset$
$A \cup A^c = S$, $A \cap A^c = \emptyset$

Disjoint Events: A and B are *disjoint* or *mutually exclusive* if A and B have no outcomes in common. This happens only if $A \cap B = \emptyset$. A collection $A_1, \ldots, A_n$ is a collection of disjoint evens if and only if $A_i \cap A_j = \emptyset, \forall i, j, i \neq j$

$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c$
$(A \cup B)^c = A^c \cap B^c$
$x \in (A \cap B)^c \Rightarrow x \notin A$ and $x \notin B$
$\Rightarrow x \in A^c$ and $x \in B^c$
$\Rightarrow x \in A^c \cap B^c$

Probabilities:

$\forall A : \Pr(A) \geq 0$
$\Pr(S) = 1$, $\Pr(\emptyset) = 0$
$\Pr(A^c) = 1 - \Pr(A)$
$A \subset B \Rightarrow \Pr(A) \leq \Pr(B)$
$\forall A : 0 \leq \Pr(A) \leq 1$

For every *infinite sequence* of *disjoint* events: $A_1, A_2, \ldots (A_i \in S)$:

$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$
$= \Pr(A_1 \cup A_2 \cup \cdots \cup A_n \cup \ldots)$
$= \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_n) + \ldots$
$\Pr(\bigcup_{i=1}^{n} A_i) = \Pr(\bigcup_{i=1}^{n} A_i + \bigcup_{i=n+1}^{\infty} \emptyset)$
$= \sum_{i=1}^{n} \Pr(A_i)$
$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

Finite Sample Spaces: $S := \{s_1, s_2, \ldots, s_n\}$, $\Pr(s_i) = P_i, \forall i = 1, 2, \ldots, n$, s.t. $\sum_{i=1}^{n} P_i = 1$. A sample space S with n outcomes $s_1, \ldots, s_n$ is a *simple sample space* if the probability assigned to each outcome is $\frac{1}{n}$. If A contains m outcomes then $\Pr(A) = \frac{m}{n}$.

Multiplication Rule: An experiment has k parts ($k \geq 2$) s.t. the i^{th} part has n_i possible outcomes, $i = 1, \ldots, k$, and *all possible outcomes can occur regardless of which outcomes have occurred in other parts.* S will contain vectors of the form $(u_1, u_2, \ldots, u_k)$. u_i is one of the n_i possible outcomes of part i. The total number of vectors is $n_1 \cdot n_2 \cdot \ldots \cdot n_k$.

Permutations: Given an array of n elements the first position can be filled with n different elements, the second with $n-1$, and so on. $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1 = n!$

$P_{n,k} = \frac{n!}{(n-k)!}$
$P_{n,n} = n!$

Combinations: In general we can "combine" n elements taking k at a time in $C_{n,k} = \frac{P_{n,k}}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$.

Multinomial Coefficients: Consider splitting n elements into $k (k \geq 2)$ groups in a way such that group j gets n_j elements and $\sum_{j=1}^{k} n_j = n$. The n_1 elements in the first group can be selected in $\binom{n}{n_1}$, the second in $\binom{n-n_1}{n_2}$, the third in $\binom{n-n_1-n_2}{n_3}$ and so on until we complete the k groups. Then: $\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \cdot \ldots \cdot \binom{n_k}{n_k} = \binom{n}{n_1, n_2, \ldots, n_k}$

Probability of union: If $A_1, A_2, \ldots, A_n$ are *disjoint events* then

$\Pr(A_1 \cup A_2 \cup \ldots \cup A_n) = \Pr(\bigcup_{i=1}^{n} A_i)$
$= \Pr(A_1) + \Pr(A_2) + \ldots + \Pr(A_n)$
$= \sum_{i=1}^{n} \Pr(A_i)$

If the events are not disjoint:

Two events $A_1, A_2 : \Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2)$

Three events: $A_1, A_2, A_3 : \Pr(A_1 \cup A_2 \cup A_3) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) - \Pr(A_1 \cap A_2) - \Pr(A_1 \cap A_3) - \Pr(A_2 \cap A_3) + \Pr(A_1 \cap A_2 \cap A_3)$

Conditional Probability: If A, B are events such that $\Pr(A) > 0$ and $\Pr(B) > 0$ then

$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$ and
$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$

Furthermore:

$\Pr(A \cap B) = \Pr(B|A) \cdot \Pr(A)$ and
$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$

In general: $\Pr(A_1 \cap A_2 \cap \ldots \cap A_n)$
$= \Pr(A_1) \cdot \Pr(A_2|A_1) \cdot \ldots \cdot \Pr(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1})$

Independence: A, B are independent events if $\Pr(A|B) = \Pr(A)$ and $\Pr(B|A) = \Pr(B)$. Then, if A, B are independent:

$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B) = \Pr(A) \cdot \Pr(B)$ and
$\Pr(A \cap B) = \Pr(B|A) \cdot \Pr(A) = \Pr(B) \cdot \Pr(A)$.

In general if $A_1, A_2, \ldots, A_n$ are independent: $\Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = \Pr(A_1) \cdot \Pr(A_2) \cdot \ldots \cdot \Pr(A_n)$.

Note that if $A \cap B = \emptyset$ then the two events are *not independent.*

Note that if A, B are independent then A, B^c are also independent.

Conditionally Independent: $A_1, A_2, \ldots, A_k$ are *conditionally independent* given B if, for every subset $A_{i_1}, \ldots, A_{i_m}$: $\Pr(A_{i_1} \cap \ldots \cap A_{i_m}|B) = \Pr(A_{i_1}|B) \cdot \ldots \cdot \Pr(A_{i_m}|B)$.

Partitions: Let $B_1, \ldots, B_k$ be such that $B_i \cap B_j = \emptyset \forall i \neq j$ and $\bigcup_{i=1}^{k} B_i = S$. Then these events form a partition of S.

$A = A \cap S = A \cap \left(\bigcup_{i=1}^{k} B_i\right)$
$= (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k)$. Then:
$\Pr(A) = \Pr(A \cap S)$
$= \Pr(A \cap \left(\bigcup_{i=1}^{k} B_i\right))$
$= \Pr(A \cap B_1) + \Pr(A \cap B_2) + \ldots + \Pr(A \cap B_k)$
$= \Pr(A|B_1) \cdot \Pr(B_1) + \Pr(A|B_2) \cdot \Pr(B_2) + \ldots + \Pr(A|B_k) \cdot \Pr(B_k)$
$= \sum_{i=1}^{k} \Pr(A|B_i) \cdot \Pr(B_i)$.

So, if $B_1, \ldots, B_k$ are a partition of S: $\Pr(A) = \sum_{i=1}^{k} \Pr(A|B_i) \cdot \Pr(B_i)$

Bayes' Theorem: Let $B_1, \ldots, B_k$:= a partition of S such that $\Pr(B_j) > 0, j \in 1, \ldots, k$. Assume you have A such that $\Pr(A) > 0$. Then: $\Pr(B_i|A) = \frac{\Pr(A|B_i) \cdot \Pr(B_i)}{\Pr(A)} = \frac{\Pr(A|B_i) \cdot \Pr(B_i)}{\sum_{j=1}^{k} \Pr(A|B_j) \cdot \Pr(B_j)}$

Random Variables: A real-valued function on S is a random variable. A random variable X is a functions that assigns a real number $X(s) = x$ to each possible outcome $s \in S$: $X : S \Rightarrow \mathcal{D}$. x := a realization of the random variable, $x \in \mathcal{D}$. We will be computing $\Pr(X \in E)$ for $E \subset \mathcal{D} = \Pr(s \in S : X(s) \in E)$

Discrete Probability Distributions: A r.v. X has a discrete distribution if it takes a countable number of values. The probability function of a discrete r.v. is $f_X(x) = \Pr(X = x)$. Properties:

$0 \leq f_X(x) \leq 1$
$\forall x \notin \mathcal{D} : f_X(x) = 0$
$\sum_{x \in \mathcal{D}} f_X(x) = 1$
$\Pr(X \in A) = \sum_{x \in A} f_X(x)$

Uniform Distribution: $X = x, x \in \{1, 2, \ldots, k\}$ with all values x equally likely. The p.f. is $f_X(x) =$

$$\Pr(X = x) = \begin{cases} \frac{1}{k} & x = 1, 2, \ldots, k \\ 0 & o.w. \end{cases}$$

Bernoulli Distribution: An event A happens with probability p:

$$X = \begin{cases} 1 & \text{if } A \text{ happens} \\ 0 & \text{if } A^c \text{ happens} \end{cases}$$

$$f_X(x) = \begin{cases} (1-p) & x = 0 \\ p & x = 1 \\ 0 & \text{o.w.} \end{cases}$$

Binomial: n Bernoulli trials repeated independently with probability of success p.

X := number of success in n trials.
$x \in \{0, 1, \ldots, n\}$.
$f_X(x) = \Pr(X = x)$

$$= \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, \ldots, n \\ 0 & \text{o.w.} \end{cases}$$

$\to X \sim Bin(n, p)$

Quantile Function: X continuous r.v. $F^{-1}(p)$ is the quantile function of X for $0 \leq p \leq 1. F^{-1}(p) = x \Rightarrow p = F(x)$.

Joint Continuous Distributions: Joint p.d.f. given by $f_{X,Y}(x,y) = \Pr((X,Y) \in A) = \iint\limits_A f(x,y)dxdy$. To find the joint c.d.f. just integrate.

Negative-Binomial:

We repeat Bernoulli trials until r successes are observed.

$X :=$ number of failures $= \{0, 1, \ldots\}$.

$p :=$ probability of success.

$\Pr(X = x) = \Pr(x$ failures before r successes)

$= \Pr(x$ failures and $r-1$ successes in $x+r-1$ trials)) $\cdot \Pr($one success in last trial)

$= \left[\binom{x+r-1}{x}(1-p)^x p^{r-1}\right] \cdot p = \binom{x+r-1}{x}(1-p)^x p^r$.

$f_X(x) = \begin{cases} \binom{x+r-1}{x}(1-p)^x p^r & x = 0,1,2,\ldots \\ 0 & \text{o.w.} \end{cases}$

Hypergeometric:

A box with A red balls and B blue balls.

n balls are drawn *without replacement*.

$X :=$ number of red balls.

$X \le \min(n, A)$.

$\max(n-B, 0) \le X \le \min(n, A) \to f_X(x) =$

$\begin{cases} \frac{\binom{A}{x}\cdot\binom{B}{n-x}}{\binom{A+B}{n}} & \max(n-B,0) \le x \le \min(n,A) \\ 0 & \text{o.w.} \end{cases}$

Geometric: Negative binomial with $r = 1$

$f_X(x) = \begin{cases} (1-p)^x p & x = 0, 1, \ldots \\ 0 & \text{o.w.} \end{cases}$

Poisson: Counts occurences of an event. X is a Poisson r.v. with parameter λ (intensity) if the p.f. is $f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x = 0,1,2,\ldots \\ 0 & \text{o.w.} \end{cases}$

with $\lambda > 0$.

Continuous Random Variables: A r.v. X has a cont. distribution if there is a non-negative f s.t. $\Pr(a \le X \le b) = \int_a^b f(x)dx$. f is the probability density function p.d.f. To find the normalizing constant integrate the p.d.f. over the domain, it must equal 1.

Cumulative Distribution Function: (c.d.f.) For any r.v. X the c.d.f. is given by $F(x) = \Pr(X \le x)$. Properties:

$\forall x : 0 \le F(x) \le 1$

$F(x)$ is non-decreasing, i.e. if $x_1 < x_2 \Rightarrow \{X \le x_1\} \subset \{X \le x_2\}$ and so $\Pr(X \le x_1) \le \Pr(X \le x_2) \Rightarrow F(x_1) \le F(x_2)$

$\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$

For a continuous r.v.:

$F(x) = \Pr(X \le x) = \int_{-\infty}^{\infty} f(t)dt$

$F'(x) = f(x)$

$\Pr(a < X \le b) = \Pr(a \le X \le b) = \Pr(a \le X < b) = \Pr(a < X < b)$

In general, $X \sim Unif[a,b] \Rightarrow f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{o.w.} \end{cases}$.

The c.d.f.: $F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1x > b \end{cases}$

Marginal Distributions: In general for discrete r.v. $f_X(x) = \sum_y f(x,y)$ and $f_Y(y) = \sum_x f(x,y)$. In the case of 2 cont. r.v. $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$. $f_X(x)$ is the marginal p.d.f. of X. $f_Y(y)$ is the marginal p.d.f. of Y.

Independence: Two r.v. are independent if they produce independent events: $\Pr(X \in A, Y \in B) = \Pr(X \in A)\cdot\Pr(Y \in B)$. This implies: $\Pr(X \le x, Y \le y) = \Pr(X \le x)\cdot\Pr(Y \le y) \Rightarrow F(x,y) = F_X(x)\cdot F_Y(y)$.

Conditional Distributions: X, Y discrete r.v. with p.f. $f_X(x), f_Y(y)$ and joint p.f. $f(x,y)$. Then:

$\Pr(X = x | Y = y) = \frac{\Pr(X=x, Y=y)}{\Pr(Y=y)} = \frac{f(x,y)}{f_Y(y)}$.

This is a new distribution and the p.f. is (p.f. of $(X|Y)$):

$g_X(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & \forall x, y : f_Y(y) > 0 \\ 0 & \text{o.w.} \end{cases}$

Note that

$\sum_x y_X(x|y) = \sum_x \frac{f(x,y)}{f_Y(y)} = \frac{1}{f_Y(y)} \cdot \sum_x f(x,y) = \frac{f_Y(y)}{f_Y(y)} = 1$.

We can also define the conditional distribution of Y given $X = x$ by:

$g_Y(y|x) = \begin{cases} \frac{f(x,y)}{f_X(x)} & \forall x, y : f_X(x) > 0 \\ 0 & \text{o.w.} \end{cases}$

In the continuous case X, Y with joint p.d.f. $f(x,y)$ and marginal p.d.f.'s $f_X(x)$ and $f_Y(y)$:

$g_X(x|y) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & f_Y(y) > 0 \\ 0 & \text{o.w.} \end{cases}$

$g_Y(y|x) = \begin{cases} \frac{f(x,y)}{f_X(x)} & f_X(x) > 0 \\ 0 & \text{o.w.} \end{cases}$.

Again note that

$\int\limits_{-\infty}^{\infty} g_X(x|y)dx = \int\limits_{-\infty}^{\infty} \frac{f(x,y)}{f_Y(y)}dx = \frac{1}{f_Y(y)} \int\limits_{-\infty}^{\infty} f(x,y)dx$

$= \frac{1}{f_Y(y)} \cdot f_Y(y) = 1$.

Multivariate Distributions: $X_1, X_2, \ldots, X_n$ have a joint discrete distribution if $(X, \ldots, X_n)$ can have only a countable sequence of values in $\mathcal{R}^n$. The joint p.f. is $f(x_1, \ldots, x_n) = \Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$. $X_1, X_2, \ldots, X_n$ have a joint continuous distribution if there exists f such that $f((X_1, \ldots, X_n) \in \mathcal{C}) = \int\limits_{\mathcal{C}} \cdots \int f(x_1, \ldots, x_n)dx_1 \ldots dx_n$. Here $f(x_1, \ldots, x_n)$ is the joint p.d.f. and $f(x_1, \ldots, x_n) = \frac{d^n F(x_1, \ldots, x_n)}{dx_1 \ldots dx_2}$. $F(x_1, \ldots, x_n) = \Pr(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$.

Marginal Distributions: $X_1, \ldots, X_n$ are cont. r.v. with joint p.d.f. $f(x_1, \ldots, x_n)$. Then the marginal distribution of $X_1 = f_{X_1}(x_1) = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)dx_2 dx_3 \ldots dx_n$.

$F(x_1)$ is the marginal c.d.f. of X_1 and $F(x_1) = \Pr(X_1 \le x_1) = \Pr(X_1 \le x_1, X_2 < \infty, \ldots, X_n < \infty)$.

Conditional Distribution: $f(x_1, \ldots, x_n)$: joint p.d.f. of $x_1, \ldots, x_n$; $f_0(x_1, \ldots, x_k)$: joint p.d.f. of $x_1, \ldots, x_k$ with $k < n$. Then $\forall x_1, \ldots, x_k$ such that $f_0(x_1, \ldots, x_k) > 0$ the conditional p.d.f. of $X_{k+1}, \ldots, X_n$ given $X_1 = x_1, \ldots, X_k = x_k$ is $g(x_{k+1}, \ldots, x_n | x_1, \ldots, x_k) = \frac{f(x_1, \ldots, x_n)}{f_0(x_1, \ldots, x_k)}$.

Functions of one R.V: Consider a r.v. X cont. with p.d.f. $f_X(x)$. Assume we are interested in $Y = r(X)$ with r a function. What is the dist. of Y? Let $G_Y(y) = \Pr(Y \le y)$ be the c.d.f. of Y: $G_Y(y) = \Pr(Y \le y) = \Pr(r(X) \le y) = \int f(x)dx\{x : r(x) \le y\}$. To get the p.d.f. of Y take derivatives: $g_Y(y) = \frac{dG_Y(y)}{dy} = \frac{1}{2}y^{-\frac{1}{2}}$. For continuous r.v. such that $Y = r(X)$ with r differentiable and one-to-one: $g_Y(y) = f_X(r^{-1}(y))\left|\frac{d}{dy}r^{-1}(y)\right|$.

Functions of two or more R.V.s: Discrete case: $X_1, X_2, \ldots, X_n$ r.v. with joint p.f. $f(x_1, \ldots, x_n$: $Y_1 = r_1(X_1, \ldots, X_n) \ldots Y_m = r_m(X_1, \ldots, X_n)$. Let $A := \{(x_1, \ldots, x_n) :=$ such that $y_1 = r(x_1, \ldots, x_n) \ldots y_m = r(x_1, \ldots, x_n)\}$. Then: $g(y_1, \ldots, y_m) = \Pr(Y_1 = y_1, \ldots, Y_m = y_m) = \sum\limits_{(x_1, \ldots, x_n) \in A} f(x_1, \ldots, x_n)$

Continuous case: $X_1, X_2, \ldots, X_n$ cont. r.v. with joint p.d.f. $f(x_1, \ldots, x_n)$. Let $Y = r(X_1, \ldots, X_n) \to A_y = \{(x_1, \ldots, x_n)$ s.t. $r(x_1, \ldots, x_n) \le y\}$ Then the c.d.f. of Y is $G(y) = \Pr(Y \le y) = \Pr(r(X_1, \ldots, X_n) \le y) = \int\limits_{A_y} \cdots \int f(x_1, \ldots, x_n)dx_1 dx_2 \ldots dx_n$. The density/p.d.f. of Y is $g(y) = \frac{dG(y)}{dy}$.

Transformations: $X_1, \ldots, X_n$ cont. r.v.'s with joint pdf $f(x_1, \ldots, x_n)$. Let $Y_1 = r_1(X_1, \ldots, X_n), \ldots, Y_n = r_n(X_1, \ldots, X_n)$. To find the joint pdf of $Y_1, \ldots, Y_2$ for a one-to-one differentiable transformation $x_1 = s_1(y_1, \ldots, y_n), \ldots, x_n = s_n(y_1, \ldots, y_n) \to$ The joint pdf of $Y_1, \ldots, Y_n$ is $g(y_1, \ldots, y_n) = f(s_1, \ldots, s_n)|J|$ where $J = \det \begin{bmatrix} \frac{ds_1}{dy_1} & \cdots & \frac{ds_1}{dy_n} \\ \vdots & \ddots & \vdots \\ \frac{ds_n}{dy_1} & \cdots & \frac{ds_n}{dy_n} \end{bmatrix}$

Linear Transformations: Suppose that $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ and $\vec{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = A\vec{X}$ (with A a non-singular matrix). Then $\vec{X} = A^{-1}\vec{Y}$ and $g_Y(y) = f_X(A^{-1}) \cdot \frac{1}{|\det A|}$

Markov Chains: A sequence of r.v.'s $X_1, X_2, \ldots$ is a stochastic process with discrete time parameter. X_1 is the initial state and X_n is the state at time n. A stochastic process with discrete time parameter is a Markov chain if for each n, $\Pr(X_{n+1} \le b | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} =\le b | X_n = x_n)$. A Markov Chain is finite if there are finite possible states. Then: $\Pr(X_1 = x_1, \ldots, X_n = x_n) = \Pr(X_1 = x_1)\Pr(X_2 = x_2 | X_1 = x_1) \cdots \Pr(X_n = x_n | X_{n-1} = x_{n-1})$. Transition distributions: When a MC has k possible states then it has a transition distribution where there exist probabilities p_{ij} for $i, j = 1, \ldots, k$ such that $\forall n : \Pr(X_{n+1} = j | X_n = i) = p_{ij}$ and if $\Pr(X_{n+1} = j | X_n = i) = p_{ij} \forall n$ then it is a stationary transition distribution. In this case there is a matrix s.t. $\sum\limits_{j=1}^{k} p_{ij} = 1, \forall i$:

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix}$$

Transition of several steps: $P^m = P \cdot \ldots \cdot P$ Just exponentiate P and then find the resulting p_{ij}.

Expectation:

$E(X) = \int\limits_{-\infty}^{\infty} xf(x)dx$

$E(X) = \sum\limits_x xf(x) = \sum\limits_x \Pr(X = x)$

If $Y = r(X)$ and $f(x)$ is the p.d.f. of X : $E(Y) = \int\limits_{-\infty}^{\infty} r(x)f(x)dx$

$Y = aX + b \to E(Y) = aE(X) + b$

a constant s.t. $\Pr(X \geq a) = 1$ then $E(X) \geq a$

b constant s.t. $\Pr(X \leq b) = 1$ then $E(X) \leq b$.

If $X_1, \ldots, X_n$ are r.v. then $E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$

$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n (E(X_i))$

$X_1, \ldots, X_n$ independent r.v.'s with finite expectation: $E\left(\Pi_{i=1}^n X_i\right) = \Pi_{i=1}^n \left(E(X_i)\right)$

Bernoulli(p): $E(X) = p$

Binomial(n, p): $E(X) = np$

Poisson: $E(X) = \lambda$

Geometric: $E(X) = \frac{1-p}{p}$

Negative Binomial: $E(X) = \frac{r(1-p)}{p}$

Variance:

$V(X) = E[(X - \mu)^x]$ with $\mu = E(X)$

S.D.: $\sigma = \sqrt{V(X)}$

$V(X) \geq 0$!!!

X discrete: $V(X) = \sum\limits_X (x - \mu)^2 f(x)$

X cont.: $V(X) = \int\limits_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$

$V(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$

$V(X) = 0 \iff \Pr(X = c) = 1$ with c constant

a, b constant: $V(aX + b) = a^2 V(X)$

$X_1, \ldots, X_n$ independent: $V(X_1 + \ldots + X_n) = V(X_1) + \ldots + V(X_n)$

Bernoulli: $V(X) = p(1 - p)$

Binomial: $V(X) = np(1 - p)$

Poisson: $V(X) = \lambda$

Geometric: $V(X) = \frac{1-p}{p^2}$

Negative Binomial: $V(X) = \frac{r(1-p)}{p^2}$

Covariance:

$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$

Discrete: $E(XY) = \sum\limits_x \sum\limits_y xyf(x, y)$

Discrete: $\mu_X = E(X) = \sum\limits_x \sum\limits_y xf(x, y)$

Discrete: $\mu_Y = E(Y) = \sum\limits_x \sum\limits_y yf(x, y)$

Cont: $E(XY) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} xyf(x, y)dxdy$

If X and Y are independent: $\text{Cov}(X, Y) = 0$

Correlation:

$\text{Corr}(X, Y) = \rho(X, Y)$

$= \frac{\text{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

$= \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$

Schwarz Inequality: $[E(UV)]^2 \leq E(U^2)E(V^2)$

Cauchy-Schwarz Ineq.: $[\text{Cov}(X, Y)]^2 \leq V(X) \cdot V(Y)$ and $-1 \leq \rho(X, Y) \leq 1$

Independent: $\text{Cov}(X, Y) = 0$ and $\rho(X, Y) = 0$

X r.v. w/ finite variance and $Y = aX + b$ s.t. $a \neq 0, a, b$, constant, then $a > 0 \to \rho(X, Y) = 1$ and $a < 0 \to \rho(X, Y) = -1$

X, Y with finite variance then $V(X + Y) = V(X) + V(Y) + 2\rho(X, Y)$

$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab\rho(X, Y)$

Other Stuff:

$\sum\limits_{i=1}^n i = \frac{n(n+1)}{2}$, $\sum\limits_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$

$\sum\limits_{i=0}^n c^i = \frac{c^{n+1}-1}{c-1}, c \neq 1$; $\sum\limits_{i=0}^\infty c^i = \frac{1}{1-c}$

$\frac{d}{dx}(a^x) = \ln a$

$(fg)' = f'g + fg'$

$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$

$\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$

$\frac{d}{dx}(e^{g(x)}) = g'(x)e^{g(x)}$

$\frac{d}{dx}(\ln g(x)) = \frac{g'(x)}{g(x)}$

$\int \frac{1}{x}dx = \ln|x| + c$, $\int \frac{1}{ax+b}dx = \frac{1}{a}\ln|ax + b| + c$

$\int e^u du = e^u + c$, $\int a^u du = \frac{a^u}{\ln a} + c$

$\int_a^b f(g(x))g'(x)dx \Rightarrow u = g(x) \Rightarrow \int_{g(a)}^{g(b)} f(u)du$

$\int udv = uv - \int vdu$