Terminology:

Sample space: The set of all possible outcomes of an experiment, denoted by Ω or S.
Symmetric sample space: Ω is symmetric if the probabilities of all $\omega \in \Omega$ are the same.
Intersection of sets: Let A and B be two events of sample space Ω. The set of all outcomes that are both in A and B is called the intersection, denoted by $A \cap B$. **Union of sets:** Set of all outcomes that are either in A or B is called the union, denoted by $A \cup B$. **Complement of set.** Set of all outcomes that are not in A but are in Ω is called the complement. It is denoted by $\bar{A}$ or A^C.
Mutually exclusive events: Two events A and B are said to be mutually exclusive if $A \cap B = \emptyset$

Basic Laws:

Law 1: Commutative laws:	$A \cup B = B \cup A \quad A \cap B = B \cap A$
Law 2: Associative laws:	$(A \cup B) \cup C = A \cup (B \cup C) \quad (A \cap B) \cap C = A \cap (B \cap C)$
Law 3: Distributive Laws:	$(A \cup B) \cap C = (A \cap B) \cup (B \cap C) \quad (A \cap B) \cup C = (A \cup B) \cap (B \cup C)$
Law 4: De Morgan's Laws:	$\overline{A_1 \cup \ldots \cup A_n} = \overline{A_1} \cap \ldots \cap \overline{A_n} \quad \overline{A_1 \cap \ldots \cap A_n} = \overline{A_1} \cup \ldots \cup \overline{A_n}$

Binomial Coefficient: $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \prod_{i=1}^{k} \frac{n+1-i}{i}$. Can also be found by using Pascal's triangle, n indicates the row (starting from 0), r refers to the element.

Axioms of probability:

Probability: Probability of E is a function satisfying three basic axioms, denoted by $P(E)$.

Axiom 1: $0 \le P(E) \le 1$	**Axiom 2:** $P(\Omega) = 1$,i.e. total probabilities must add up to 1.
Axiom 3: For any sequence of mutually exclusive events $A_1, \ldots, : P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$	
Theorem 1: $P(\emptyset) = 0$	**Theorem 2:** For a finite collection of mutually exclusive events $A_1, \ldots, A_n$, $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
Theorem 3: $P(\bar{A}) = 1 - P(A)$	**Theorem 4:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Basics of Combinatorics:

Case 1	**Order:** $+$ (permutation) **Repetition:** $+$ **Identity:** n^k (choose from n objects k times)	Number of options to choose k objects out of n, when the order of selection is relevant with repetition. You make k different selection (because there are repetitions). Each time you have n options to choose from (forming a word from letters, phone numbers)
Case 2	**Order:** $+$ (permutation) **Repetition:** $-$ **Identity:** Row: $n!$, Circle: $(n-1)!$	Number of permutations to arrange n objects in a row is $n!$. Number of permutations to arrange n objects in a circle is $(n-1)!$ (because the first object has one option)
Case 3	**Order:** : $-$ (combination) **Repetition:** : $-$ **Identity:** $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ (choose k objects out of n)	Number of options to choose k objects out of n without repetition and when the order of selection is irrelevant. You choose all n in a row that you divide by the number of internal arrangements for both the k selected objects and the $(n-k)$ non-selected objects (order is irrelevant)

Conditional Probability:

For two events A and B in sample space Ω, where $P(B) > 0$, the conditional probability, i.e. the probability that A will occur after B has already occurred is defined as: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

X, given $Y = y$	**Conditional Probability**	**Conditional Expectation**	**Conditional Variance**
Discrete	$P_{X\|Y}(x\|y) = \frac{P(X=x, Y=y)}{P(Y=y)}$	$E(X\|Y=y) = \sum_x x P_{X\|Y}(x\|y)$	$V(X\|Y) = E[X - E[X\|Y])^2\|Y]$
Continuous	$f_{X\|Y}(x\|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$	$E(X\|Y=y) = \int_{-\infty}^{\infty} x f_{X\|Y}(x\|y)dx$	$V(X\|Y) = E[X^2\|Y] - E[X\|Y]^2$

Theorem 7: $P(A_1 \cap \ldots \cap A_n) = \prod_{k=1}^{n} P(A_k | \cap_{l=1}^{k-1} A_l)$

Bayes' Theorem:

Division: A set of events $A_1, \ldots, A_n$ is called a division of the sample space Ω, if $\bigcup_{k=1}^{n} A_k = \Omega$, and $A_i \cup A_j = \emptyset$ for $i \neq j$
Bayes' Theorem: Given a division $A_1, \ldots, A_n$ of the sample space Ω, and an event B in Ω, then $P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$

Independent events:

Two independent events: Two events, A and B are said to be independent if $P(A \cap B) = P(A)P(B)$
Theorem 9: $P(A|B) = P(A)$ if and only if A and B are independent.
Alternative independence formula: $P(EF) = P(E)P(F)$ hold for independent events.
Theorem 10: If A and B are independent, then so are $\bar{A}$ and B, A and $\bar{B}$, $\bar{A}$ and $\bar{B}$

Discrete Random Variables:

Random variable: Function $X: \Omega \to \mathbb{R}$, maps points from the sample space to the real line is called a random variable.
Discrete Random Variable: An RV that can have at most a countable number of possible values is said to be discrete.

Continuous Random Variables:

Continuous CDF: The CDF of a continuous RV is a continuous function (not only right continuous).
Continuous RV Probability Density: $P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. properties: $f_X(x) \ge 0$, and $\int_{-\infty}^{\infty} f_X(t)dt = 1$
Calculating probabilities with PDF: $P(a \le X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$. We let $a = b$ in this equation, we get $P(X = a) = \int_a^a f(x)dx = 0$. This states that the probability that a continuous RV will assume any fixed value is zero.
Expectation and variance: $E[X] = \int_{-\infty}^{\infty} x f(x)dx$, $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx, V(X) = E[(X - E[X])^2]$

Probability Mass Function:

Probability mass function: A function which gives the probability of a discrete random variable X having value x is called the probability mass function of X. It is denoted as $P(X = x)$
Expectation value: If X is a discrete random variable having a probability mass function $P(X = x)$, then the expectation or the expected value of X is defined to be $E[X] = \sum_{\{x|P(X=x)>0\}} xP(X = x)$
Theorem 11: If X is a discrete random variable that takes on one of the values x_i where $i \in \mathbb{N}$, with probability mass function $P(X = x_i)$, then for any real valued function g, $F[g(x)] = \sum_i g(x_i)P(x = x_i)$

Variance:

Variance: The variance of a random variable X is defined to be $V(x) = E[(X - E[X])^2] = E[X^2] - E[X]^2$
$V(X)$ using conditional variances: $V(X) = E[V(X|Y) + V(E[X|Y])$
Theorem 12: $V(aX + b) = a^2 V(x)$
Continuous Random Variable: A random variable X is said to be a continuous random variable if there exists a function f such that $F_X(x) = \int_{-\infty}^{x} f(t)dt$
Probability Density Function: Let $F_x(x) = \int_{-\infty}^{x} f(t)dt$, then $f(t)$ is the probability density function of x

Bernoulli and Binomial Random Variables:

Bernoulli random variable: (Model: Probability of Success in a single trial) A random variable X is said to be a Bernoulli random variable if its probability mass function is given by $P(X = 0) = 1 - p$, $P(X = 1) = p$. Denote Bernoulli variables as $X \sim \mathrm{Ber}(p)$
Expectation and Variance: For a Bernoulli random variable, $E[x] = p$, $V(x) = p(1 - p)$
Binomial random variable: Consider n independent trials, each of which has a probability of success p, and the probability of failure $1 - p$. If X represents the number of successes that occur in the n trials, then X is said to be a binomial random variable with parameters (n, p). It is denoted as $X \sim Bin(n, p)$
Binomial Probability Mass: For a binomial rand. variable $X \sim \mathrm{Bin}(n, p)$ the probability mass is

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$$

$$\boxed{\begin{aligned} E[x] &= np \\ V[x] &= np(1-p) \end{aligned}}$$

Expectation and Variance: Let $X \sim Bin(n, p)$ and $Y \sim Bin(n - 1, p)$, then $E[X^k] = npE[(Y + 1)^{k-1}]$

Poisson Random Variables:

Poisson Random Variables: A random variable X that takes on whole number values is said to be a Poisson random variable with parameter λ if for some $\lambda > 0$, $P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!}$,where $i \ge 0 \in \mathbb{W}$.
It is denoted as $X \sim Poi(p)$. Examples of Poisson Variables: The number of people in a population who survive to age 100, the number of customers entering a post office on a given day, the number of wrong phone numbers dialed in a day, etc.

Theorem 18: Let $X \sim Bin(n,p)$. If $n \to \infty$, the probability distribution of X is a Poisson distribution.
Expectation and Variance: Let $X \sim Poi(\lambda)$, then $E[X] = \lambda$ and $V[X] = \lambda$
Assumptions for Poisson Distributions for Events over a Period of Time:

1. The probability that an event occurs in an interval of length h is $\lambda h + o(h)$
2. For a small enough h, the probability that two or more events occur in an interval of length h s small, i.e. is $o(h)$
3. The number of events in intervals that are not overlapping are independent.

Properties of Poisson Distribution:

1. If the number of events in a time/area unit is $\sim \mathrm{Poi}(\lambda)$, then the number of events in n units of time/areas $\sim \mathrm{Poi}(n\lambda)$ (i.e. the sum of independently and identically distributed Poisson random variables is a Poisson random variable with a rate of the sum of rates, $n\lambda$
2. In 2 disjoint time/area segments, the number of events is independent.
3. Two events cannot happen at the same time.

Poisson Approximation of the Binomial: If $n \to \infty, p \to 0$ (i.e the number of trials is very large, and the probability of success is very low) then: $\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!}$. In practical terms, we can use this approximation when $np < 5 \sim \mathrm{Poi}(np = \lambda)$

Geometric Random Variables:

Geometric Random Variable: Suppose that independent trials, each having probability of success $0 < p < 1$ are performed until a success occurs. If X is the number of trials required, then X is said to have a geometric distribution. It is denoted as $X \sim Geo(p)$. The probability distribution of X is $P(X = n) = (1-p)^{n-1} p$
Expectation and Variance: Let as $X \sim Geo(p)$, then $E[x] = \frac{1}{p}$ and $V(X) = \frac{1-p}{p^2}$
Properties: (1) Lack of memory $\to P(X = n + k | X > n) = P(X = k)$ **(2)** $P(X > n) = (1-p)^n$

Negative Binomial Random Variable:

Negative Binomial Random Variable: Suppose that independent trials, each having probability of success $0 < p < 1$ are performed until a total of r successes are accumulated. If X is the number of trials required, then X is said to be a negative binomial random variable. It is denoted as $X \sim NB(r,p)$. The last trial must necessarily result in a success, and there must be $r - 1$ more successes in the first $n - 1$ trials. Therefore, the probability distribution of X is $P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$
Theorem 23: $n \binom{n-1}{r-1} = r \binom{n}{r}$
Theorem 24: Let $X \sim NB(r,p)$, then $E[X] = \frac{r}{p}$ and $V(X) = \frac{r(1-p)}{p^2}$

Hypergeometric Random Variable:

Hypergeometric random variable: Suppose that a sample of size n is to be chosen randomly and without replacement from a population of N, of which m posess a particular characteristic, and the other $N - m$ do not. If X is the number of individuals in the selected sample, then X is said to be a Hypergeometric distribution. It is denoted as $X \sim HG(n,N,m)$.
The probability distribution of X is $P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$
Expectation and Variance: Let $X \sim HG(n,N,m)$, then $E[X] = \frac{nm}{N}$ and $V(X) = n\frac{m}{N}\left(1 - \frac{m}{N}\right)\left(\frac{N-n}{N-1}\right)$

Uniform Random Variable

Uniform Random Variable: A random selection of an integer between a and b. Each value is equally likely to be observed. A random variable X is said to be a uniform random variable over the interval (a,b) if its probability density function (continuous) is $f(x) = \begin{cases} \frac{1}{b-a} ; a < x < b \\ 0 ; \text{otherwise} \end{cases}$. For a Discrete Uniform Random Variable, the probability Mass Function is $P(X = k) = \frac{1}{n}$ It is denoted as $X \sim U(a,b)$

Uniform CDF: The CDF of a uniform random variable X is $F_X(x) = \begin{cases} 0 ; x < a \\ \frac{x-a}{b-a} ; a \le x \le b \\ 1 ; b < x \end{cases}$
Expectation and Variance: Let $X \sim U(a,b)$, then $E[X] = \frac{a+b}{2}$ and $V(X) = \frac{(b-a)^2}{12}$ $(V_{discrete}(x) = \frac{(b-a+1)^2 - 1}{12}$

Exponential Random Variable

Exponential Random Variable: Given a Poisson process, with a rate of λ, where X is the duration until the first event, or the duration between the 1st and 2nd events, etc (E.g. the time between 2 clients entering the bank), then $X \sim \mathrm{Exp}(\lambda)$ A random variable X is said to be an exponential random variable over the interval (a,b) if its probability density function is $f(x) = \begin{cases} \lambda e^{-\lambda x} ; x \ge 0 \\ 0 ; x < 0 \end{cases}$. It is denoted as $X \sim \mathrm{Exp}(\lambda)$
Cumulative Distribution Function: The cumulative distribution function of an exponential random variable X with parameter λ is $F_X(x) = \begin{cases} 1 - e^{-\lambda x} ; x \ge 0 \\ 0 ; x < 0 \end{cases}$
Expectation and Variance: Let $X \sim Exp(\lambda)$, then $E[X] = \frac{1}{\lambda}$; $E[X^n] = \frac{n!}{\lambda^n}$; $V(X) = \frac{1}{\lambda^2}$
Properties: Memory-less - $P(X > s + t | X > t) = P(X > s), \forall s,t \ge 0$. In other words, $\frac{P(X > s+t, X > t)}{P(X > t)} = P(X > s)$ or $P(X > s + t) = P(X > s)P(X > t)$. For example, if we think of X as being the lifetime of some instrument, the equation states that the probability that the instrument survives for at least $s + t$ hours, given that it has survived t hours, is the same as the initial probability that it survives for at least s hours. In other words, if the instrument is alive at age t, the distribution of the remaining amount of time that it survives is the same as the original lifetime distribution. That is, it is as if the instrument does not "remember" that is has already been in use for a time t.

Normal Random Variable

Normal Random Variable: A random variable X is said to be a normal random variable if its probability density function is $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where μ and σ^2 are parameters. If $\mu = 0, \sigma^2 = 1$ then X is said to be a standard normal random variable.
Cumulative Distribution Function: CDF of a standard normal X $(X \sim N(0,1)$ is denoted as $\Phi(x)$. It is defined as $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$
CDF Property: Let X be a standart normal random variable, then: $\Phi(-x) = 1 - \Phi(x)$
Theorem 36: If X is normally distributed with parameters μ and σ^2, then $Y = ax + b$, where $a > 0$ is normally distributed with parameters $a\mu + b$ and $a^2\sigma^2$
Theorem 37: If X is normally distributed with parameters μ and σ^2, then $Z = \frac{X-\mu}{\sigma}$ is normally distributed with parameters 0 and 1
Standard Expectation and Variance: Let Z be a standart normal random variable, then $E[X] = 0, V(X) = 1$
General Expectation and Variance: Let X be a non-standard normal random variable, then $E[X] = \mu$, and $V(X) = \sigma^2$

De Moivre-Laplace Limit Theorem

De Moivre-Laplace Limit Theorem: Let S_n be the number of successes that occur when n independent trials, each with probability of success p, are performed. Then, for any $a < b$, $\lim_{n\to\infty}\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} < b\right) = \Phi(b) - \Phi(a)$

Combination of Random Variables:

Theorem 43: Expected value of sums of random variables: Let $X = \sum_{i=1}^{n} X_i$ where each X_i is a random variable, possibly of different distributions. Then, $E[X] = \sum_{i=1}^{n} E[X_i]$

Joint Cumulative Probability Distribution Function:

Joint Cumulative PDF: For any two random variables X and Y, the joint cumulative probability distribution function of X, Y is defined to be $F_{X,Y}(a,b) = P(X \le a, Y \le b)$, where $a, b \in \mathbb{R}$

Joint Continuous Variables:

Independence: Two RV are independent if for *every* pair of outcomes x, y: $P(X = x, Y = y) = P(X = x)P(Y = y)$

Joint continuity: X and Y are said to be jointly continuous if there exists a function $f(x,y)$ defined for all real x and y, such that for every set C of pairs of real numbers, $P\big((x,y)\in C\big)=\iint_{(x,y)\in C} f(x,y)dxdy$.

Joint PDF: The function $f(x,y)$ is called the joint probability density function of X and Y. The joint cPDF is $F_{X,Y}(a,b)=P(-\infty\le X\le a,-\infty\le Y\le b)=\int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)dxdy$. Therefore: $f(a,b)=\frac{\partial^2}{\partial a\partial b}F(a,b)$

Marginal Densities: $f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy$, $f_Y(y)=\int_{-\infty}^{\infty} f(x,y)dx$

Function of Independent Variables – 3 stage method:

1. Draw the support (all possible values) of X,Y in the coordinate system
2. Draw $g(s,t)$ →Determine D (integration borders)
3. $\iint_D f_{X,Y}(s,t)d(s,t)=P(g(X,Y)\le C)$

Independent Random Variables:

Independent random variables: Two random variables X and Y are said to be independent if $P(X\in A,Y\in B)=P(X\in A)P(Y\in B)$, where $A,B\subseteq\mathbb{R}$

Theorem 45: X and Y are independent random variables if and only if $F_{X,Y}(a,b)=F_X(a)F_Y(a)$

Theorem 46: If X and Y are discrete, then $P(X=x,Y=y)=P(X=x)P(Y=y)$ for all (x,y), if and only if X and Y are independent.

Theorem 47: If X and Y are continuous, then $f(x,y)=f_X(x)f_Y(y)$ for all (x,y), if and only if X and Y are independent.

Theorem 48: IF X_i, for $i\in\mathbb{N}$ are independent normal variables, with parameters μ_i and σ_i^2 respectively, then $\sum_{i=1}^{n} x_i$ is a normal random variable with parameters $\sum_{i=1}^{n}\mu_i$ and $\sum_{i=1}^{n}\sigma_i^2$

Theorem 49: Let $X=Poi(\lambda_1)$, $Y=Poi(\lambda_2)$ be independent random variables. Then, $X+Y\sim Poi(\lambda_1+\lambda_2)$

Theorem 50: Let $X\sim Bin(n,p)$, $Y\sim Bin(m,p)$ be independent random variables, then $X+Y\sim Bin(n+m,p)$

Theorem 51: Let $X_i\sim Geo(p)$ be independent random variables, for $i\in\mathbb{N}$. Then, $\sum_{i=1}^{n} X_i=NB(n,p)$

Conditional Random Variables:

Theorem 52: The conditional PMF of two discrete random variables X and Y is $P(X=x|Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}$

Theorem 53: The conditional probability density function of two continuous random variables X and Y (alt. the cond. Distribution of X given $Y=y$) is $f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$, and the cumul. distribution function is $F_{X|Y}(a,y)=\int_{-\infty}^{a} f_{X|Y}(x|y)dx$.

$P(X>1|Y=y)=\int_{1}^{\infty} f_{x|y}(x|y)$ | $f_{x|y}(x|y)=\frac{f(x,y)}{f_Y(y)}$ | $f_y(y)=\int_{-\infty}^{\infty} f(x,y)dx$

Properties of Expectation:

Let X be a continuous random variable. Then, the expectation is defined as $E[X]=\int_{-\infty}^{\infty} xf(x)dx$

Theorem 14: If X is a continuous random variable, then for any real function g, $E[g(x)]=\int_{-\infty}^{\infty} g(x)f(x)dx$

Theorem 54: If X and Y have a joint probability mass function $P(X=x,Y=y)$, then $E[g(X,Y)]=\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty} g(x,y)P(X=x,Y=y)$

Theorem 55: If X and Y have a joint probability density function $f(x,y)$, then $E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)$

Sample: Let $X_1,\dots,X_n$ be independent and identically distributed random variables, having distribution function F, and expected value μ. Such a sequence of random variables is said to constitute a sample from the distribution F.

Sample mean: Let $X_1,\dots,X_n$ be a sample with F *and* μ. Then the sample mean of the sample is : $\bar{X}=\sum_{i=1}^{n}\frac{X_i}{n}$

Theorem 56: Let $X_1,\dots,X_n$ be a sample with F *and* μ. Then $E[\bar{X}]=\mu$

Theorem 59: If X and Y are independent, then, for any function g and h, $E[g(X)g(Y)]=E[g(X)]E[h(Y)]$

Theorem 60: Let X and Y be jointly continuous with joint density function $f(x,y)$, then: $E[g(X)g(Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)dxdy$. If X and Y are independent, refer to Theorem 59.

Conditional Expectation: Let X and Y be two random variables. The conditional expectation of X given $Y=y$ is the weighted average of all the values X can take, weighted by the respective conditional probability. Let $P_{X|Y=y}(x)$ be the conditional PMF of X given Y-y. The conditional expectation is given by $E[X|Y=y]=\sum_{all\ x} xp_{X|Y=y}(x)$

Properties of Variance:

Uncorrelated random variables: X and Y are said to be uncorrelated if and only if $\mathrm{Cov}(X,Y)=0$

Sample variance: Let $X_1,\dots,X_n$ be a sample with expected value μ and variation σ^2. Then the following is called the sample variance of the sample: $S^2=\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{n-1}$

Covariance and Correlation:

Covariance: Let X and Y be random variables. The covariance quantifies correlation between 2 variables. If it is positive, there is positive relationship between the 2 variables (i.e. when one increases (decreases) the other increases (decreases)), and when it is negative, they are inversely correlated (i.e. when one increases (decreases) the other decreases (increases)). Covariance is defined as follows: $\mathrm{Cov}(X,Y)=E\big[(X-E[X])(Y-E[Y]\big]=E(XY)-E(X)E(Y)$

Special Case of Covariance: $\mathrm{Cov}(X,X)=E(X^2)-E^2(X)=Var(x)$

Properties of Covariance	
Symmetry: $\mathrm{Cov}(X,Y)$=$\mathrm{Cov}(Y,X)$	Special Case: $\mathrm{Cov}(X,X)=V(X)$
$\mathrm{Cov}(aX,Y)$=V(X)	$V(X\pm Y)=V(X)+V(Y)\pm 2\mathrm{Cov}(X,Y)$
$\mathrm{Cov}(X,Y+Z)=\mathrm{Cov}(X,Y)+\mathrm{Cov}(X,Z)$	$\mathrm{Cov}(a+bX,c+dY)=ad\mathrm{Cov}(X,Z)$
$\mathrm{Cov}\big(\sum_{i=1}^{n} X_i\,,\sum_{j=1}^{m} Y_j\big)$=$\sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{Cov}(X_i,Y_j)$	
$\mathrm{Cov}\big(\sum_{i=1}^{n}(a_iX_i+b_i)\,,\sum_{j=1}^{m}(d_jX_j+c_j)\big)=\sum_{i=1}^{n}\sum_{j=1}^{m} a_id_j\mathrm{Cov}(X_i,Y_j)$	
$V(\sum_{i=1}^{n} X_i)=\sum_{i=1}^{n} V(X_i)+2\sum_{i=1}^{n}\mathrm{Cov}(X_i,X_j)$	

Correlation: $\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)V(Y)}}$. The correlation is between $-1\le\rho\le 1$, where the limits indicate perfect linear correlation between variables: $|\rho|=1\ \Leftrightarrow X=\alpha Y+\beta$, where the sign of α equals the sign of ρ.

Independence VS. Correlation: If X and Y are independent, then $\mathrm{Cov}(X,Y)=0$. Independent RV are uncorrelated BUT uncorrelated RV are not necessarily independent.

Independence, Mean and Variance: X, Y are independent and $E(XY)=\int_0^{\infty}\int_0^{\infty} xyf_{X,Y}(x,y)dxdy=\int_0^{\infty}\int_0^{\infty} xyf_X(x)f_Y(y)dxdy=\int_0^{\infty} xf_X(x)dx\cdot\int_0^{\infty} yf_Y(y)dy$. Therefore, for each number of RV: $Var(\sum_{i=1}^{n} X_i)=\sum_{i=1}^{n} Var(X_i)+2\sum_{1\le i<j<n}\mathrm{Cov}(X_i,X_j)$. However, when the variables are uncorrelated (let alone – independent), the variance of the sum is the sum of variances (delete the last expression when uncorrelated)

Hypothesis Testing:

Parameter: A characteristic of the distribution of a population, for example μ,σ^2,λ.

Statistic: Any function of a number of random variables, usually identically distributed, that may be used to estimate a population parameter. Example: $X_1, X_1+X_3, \frac{\sum X_i}{n}=\bar{X}$

Central Limit Theorem: Let $\{X_n\}$ be a sequence of independent identically distributed random variables, with expectation μ and variance σ^2. Let $\overline{X_n}$ be the sample mean of the first n terms of the sequence. $\overline{X_n}=\frac{1}{n}\sum_{i=1}^{n} X_i$. Let the definition of the standardized sample mean be: $Z_n=\frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}$. Then, for every $t\in\mathbb{R}$: $F_{Z_n}(t)\xrightarrow{n\to\infty}\Phi(t)$. Loosely put, this states that the sum of a large number of independent random variables (identically distributed) has a distribution that is approximately normal.

1. We observe a sample consisting of n observations $X_1,X_2,\dots,X_n$
2. If n is large enough, then a standard normal distribution is a good approximation of the distribution of $\frac{\overline{X_n}-\mu}{\sigma}\sqrt{n}$
3. Therefore we pretend that $\frac{\overline{X_n}-\mu}{\sigma}\sqrt{n}\sim N(0,1)$
4. As a consequence, the distribution of the sample mean $\overline{X_n}$ is $\overline{X_n}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$

Confidence Level: Denoted by α, is the area under one/both tails. The confidence level represents the frequency (i.e. proportion) of possible confidence intervals that contain the true value of the unknown parameter. In other words, it quantifies the level of confidence that the parameter lies in the interval.

Confidence interval: Interval for (approximately) normal population (known variance), if $n\ge 25$, then according to CTL (known variance): A confidence interval for μ with $1-\alpha$ confidence level is: $\bar{X}\pm z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$. To find confidence interval, identify the sample mean $\bar{x}$. Identify whether the standard deviation σ is known or unknown, s. If standard deviation is known then $z^* = -\Phi^{-1}\left(\frac{\alpha}{2}\right)$, where Φ is the CDF of standard normal distribution. This value is only dependent on the confidence level for the test. If The standard deviation is unknown ,then t-distribution is used as the critical value. This value is dependent on the confidence level C for the test and degrees of freedom. The degrees of freedom are found by subtracting 1 from the number of observations, $n-1$. The critical value is found from the t-distribution table. In this table, the $\alpha = \frac{1-C}{2}$. Plug the found values into the appropriate equations: For a known standard deviation: $\left(\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}}\right)$. For unknown standard deviation: $\left(\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}\right)$. The final step is to interpret the answer. The confidence interval represents values for the population mean for which the difference between the mean and the observed estimate is not statistically significant at chosen level. **Half-length of the confidence interval:** $d = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

Z-TEST:

Z-test (one expectation): $Z_n = \frac{\overline{X_n} - \mu}{\sqrt{\sigma^2/n}}$, where $\bar{x_n}$ is the sample mean, μ and σ are expectation and (expected) stdev respectively, and n is the sample size. Look up a value of z on the z-table, based on your given alpha, and compare it to Z_n. If $Z_n > z \to H_0$ not valid. If $Z_n < z \to H_0$ is valid. If you're performing a 2-tailed test, extend the comparison intervals to the other tail. Usefull: $\overline{X_n} = \mu_0 + \frac{\sigma}{\sqrt{n}} z$

Z-test (proportion, crit. Value approach): Consider a parameter p of population proportion. For instance we might want to know the proportion of males within a total population of adults. A test of proportion will assess whether a sample from a population represents the true proportion from the entire population. $Z_n = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, where p_0 is the null hypothesized proportion, i.e., when $H_0: p = p_0$, and $\hat{p}$ is the sample proportion. Method: State the null and alternative hypothesis, calculate the test statistic, determine the critical region, and determine if the test statistic falls in the critical region. If not – reject the null hypothesis.

Z-test (proportion, p-value approach): State the null hypothesis and the alternative hypothesis. Determine the level of significance α (usually given). Calculate the p-value. If $p < \alpha \to H_0$ not valid. Otherwise, H_0 is valid.
Half length: $d = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ **Level of confidence:** $1 - \alpha = P(Z_{\frac{\alpha}{2}} \le Z \le Z_{1-\frac{\alpha}{2}})$

T-TEST

T-test (one sample): $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$, where S=sample stdev. Compare the calculated t with the t you look up in the t-table for a given α. Use the same conditions as for the z-test to determine hypothesis validity.
T-test (paired sample): Because half of the sample now depends on the other half, the paired version of the test has only $\frac{n}{2} - 1$ degrees of freedom.

T-test (indep. sample, eq. var., any n): $t = \left((\overline{X_1} - \overline{X_2}) - \overbrace{(\mu_1 - \mu_2)}\limits_{\text{amost always 0}}\right) \Big/ \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$.

Sample variance (t-test): $S_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$ where $SS_n = S_{tdev_n}^2 \cdot df_n$, where df_n=deg. of freedom

Confidence Interval (t-test indep sample): $(\overline{X_1} - \overline{X_2}) \pm t_{\text{from table}}\left(\sqrt{S_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}\right)$

Confidence Interval:

$$CI = \hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Hypothesis test, Normal Distribution – Unknown Variance: Suppose we have a Sample of $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ with unknown μ, σ^2. Since don't know σ^2, we will have to estimate it from the sample. In order to test hypothesis on μ, we will use the following Unbiased Estimator for σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x_n})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - n\bar{x_n}^2)$. The Test Statistic: $T_{\bar{X}} = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{n-1}$. There are 3 possible alternatives (left, right, 2-sided). We make the following Assumptions: Observations are normally distributed (not only according to CLT approximation), and Variances do not change. We reject the null if:

Right Tail: $\bar{X} > \mu_0 + t_{n-1,1-\alpha} \frac{S}{\sqrt{n}}, \frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{n-1,1-\alpha}, p - value = P(t_{n-1} \ge T_{\bar{X}}) < \alpha$

Left Tail: $\bar{X} > \mu_0 - t_{n-1,1-\alpha} \frac{S}{\sqrt{n}}, \frac{\bar{X} - \mu_0}{S/\sqrt{n}} > -t_{n-1,1-\alpha}, p - value = P(t_{n-1} \le T_{\bar{X}}) < \alpha$

2-Tail: $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > -t_{n-1,1-\frac{\alpha}{2}}$ OR $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{n-1,1-\frac{\alpha}{2}}, \bar{X} < \mu_0 - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$ OR $\bar{X} > \mu_0 + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$, p-val $= P(t_{n-1} \ge T_{\bar{X}}) < \alpha$

Paired-sample t-test: Hypothesis test on the expectations of 2 dependent populations (e.g. before-after), we will use a paired sample and test the differences. Let $X_1, \dots, X_n$ be the first sample, and $Y_1, \dots, Y_n$ the second sample. The Differences are defined as $D_i = X_i - Y_i$. We Test the Hypothesis: $H_0: \mu_X = \mu_y \Leftrightarrow \mu_D = 0, H_1: \mu_X \ne \mu_y \Leftrightarrow \mu_D \ne 0$. The Statistic: $T_0 = \frac{\overline{D_n}}{S/\sqrt{n}} \sim t_{n-1}$

Independent-sample t-test: We make two Assumptions: the observations are from normal distribution(s), and both populations have the same variance. Therefore, we use the following Estimate for Variance:
$S_p^2 = \frac{\sum_{i=1}^{n_x}(X_i - \overline{X_{n_x}})^2 + \sum_{i=1}^{n_y}(Y_i - \overline{Y_{n_y}})^2}{n_x + n_y - 2} = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$. The Standardized Statistic is: $T_0 = \frac{\bar{Y} - \bar{X}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x + n_y - 2}$. Confidence Interval: $\bar{Y} - \bar{X} \pm S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \cdot t_{n_x + n_y - 2, 1-\frac{\alpha}{2}}$

Alt. Hypothesis	
$H_1: \mu_x < \mu_y$	$C = \{T_0 > t_{n_x+n_y-2,1-\alpha}\}, p - value = P\left(t_{n_x+n_y-2} > T_0\right)$
$H_1: \mu_y < \mu_x$	$C = \{T_0 < -t_{n_x+n_y-2,1-\alpha}\}, p - value = P\left(t_{n_x+n_y-2} < T_0\right)$
$H_1: \mu_x \ne \mu_y$	$C = \{T_0 < -t_{n_x+n_y-2,1-\alpha}\} \cup \{T_0 > t_{n_x+n_y-2,1-\alpha}\}$

Type 1 Error: A type I error occurs when the null hypothesis (H_0) is true, but is rejected. The type I error rate or significance level is the probability of rejecting the null hypothesis given that it is true] It is denoted by the letter α (alpha) and is also called the alpha level.
Type 2 Error: A type II error occurs when the null hypothesis is false, but erroneously fails to be rejected. The Probability of a Type II error usually cannot be computed as it depends on the unknown population mean. However, it can be found for given values of μ, σ^2, n. Type II error $= \beta$.

Table of Error Types		Null Hypothesis H_0 is	
		True	**False**
Decision about Null Hypothesis H_0	**Not Rejected**	Correct inference	Type II Error
	Rejected	Type I Error	Correct Inference

Power: The power of a hypothesis test is equal to $1 - \beta$, where $\beta = P(\text{Type II Error})$ Essentially the power of a test is the probability that we make the right decision when the null hypothesis is not correct (i.e. we correctly reject it).
Type II Error z-test: $P(\text{Type II Error}) = P\left(z > \frac{\overline{X_{rej}} - \mu_{true}}{\sigma\sqrt{n}}\right)$, where X_{rej} is the rejection value/region expressed in terms of $\bar{X}_n$, instead of Z_n, and μ_{true} is the true value of the mean.

Nine numbered balls are drawn randomly, one by one, and without repetition from an urn. $\Omega = 9$!

a) *What is the probability that they are drawn according to their number, meaning, 1 first, then 2m etc?*
$\Omega = 9!$, answer: $\frac{1}{9!}$

b) *What is the probability that ball number 5 is drawn before ball number 2*
Answer: $\frac{1}{2}$, because of symmetry – no difference between the 2 balls, and one of them will be drawn before the other.

c) *What is the distribution of the number of balls with even numbers that are drawn within the first five draws*
X –number of balls with even numbers within 5 first draws. There are 4 even numbers: $\{2,4,6,8\}$, and 5 odd numbers $\{1,3,5,7,9\}$. Therefore, $X \in [0,4]$. $X \sim \text{HG}(k = 4, N = 9, n = 5)$. Find probabilities of X=0,1,2,3,4 using Hypergeometric formulae.

$\Omega = 9 \cdot 8 \cdot 7 = 504$.Write down all the possible cases which satisfy the requirement $P(E) = 84/504$

1st	*2nd*	*# of possible 3rd*	*Total*	*1st*	*2nd*	*# of possible 3rd*	*Total*
7	8	9	1	3	4,5,6,7,8	5,4,3,2,1	15
6	7,8	9	3	2	3,4,5,6,7,8	6,5,4,3,2,1	21
5	6,7,8	3,2,1	6	1	2,3,4,5,6,7,8,9	7,6,5,4,3,2,1,	28
4	5,6,7,8	4,3,2,1,	10	Sum total of possible satisfactory events			84

Question: A die is rolled 100 times.

a) *Let X be the number of times the result was even, and let Y be the number of times the result was 4. Find the joint probability mass function of (X,Y).* Solution: $X_{\text{even}} \sim \text{Bin}\left(n = 100, p_x = \frac{1}{2}\right)$, $Y_{="4"} \sim \text{Bin}\left(100, p_y = \frac{1}{6}\right)$.

$P(X \cap Y) = P(X = x, Y = y) = P(Y|X) \cdot P(X) = P(Y = y|X = x)P(X = x)$. $P(Y = y|X = x) = \binom{x}{y}\left(\frac{1}{3}\right)^y_{\{4\}}\left(\frac{2}{3}\right)^{x-y}_{\{2,6\}}$. $Y|X = x \sim \text{Bin}\left(X_{\{2,4,6,\}}, \frac{1}{3}\right)$. $P(X = x) = \binom{100}{x}\left(\frac{1}{2}\right)^{100}$. $(X = x, Y = y) = P(Y|X) \cdot P(X) = \overbrace{\binom{x}{y}\left(\frac{1}{3}\right)^y\left(\frac{2}{3}\right)^{x-y}}^{P(Y|X)} \cdot \overbrace{\binom{100}{x}\left(\frac{1}{2}\right)^{100}}^{P(X)}$

b) *What is the conditional expectancy of Y when it is known that $X = 2$*

By $P(Y|X)$ from prev. part: $P(Y|X = 2) = \binom{2}{y}\left(\frac{1}{3}\right)^y\left(\frac{2}{3}\right)^{2-y} \cdot \binom{100}{x}\left(\frac{1}{2}\right)^{100}$. In addition, from (a): $Y|X = 2 \sim \text{Bin}(2, \frac{1}{3})$. By the formula for Binomial Expectation: $E(Y|X = 2) = np = \frac{2}{3}$

c) *Are X and Y independent?* Answer: They are NOT independent, can be explained 2 ways:

1. Both X,Y appear together in the joint mass function, therefore can't be independent.
2. Counter example: $P(Y = 0|X = 0) \neq P(Y = 0) \cdot P(X = 0) = 1 \neq \left(\frac{1}{6}\right)^{100}\left(\frac{1}{2}\right)^{100}$

d) *Let X be the number of times the result was even, and let W be the number of times the result was smaller than 3. Are X and W independent?*

$X \sim \text{Bin}\left(100, \frac{1}{2}\right)$, even: $\{2,4,6,\}$. $W \sim \text{Bin}(100, \frac{1}{3})$: $\{1,2\}$. The events are independent. To prove: Denote a single roll, i=1,…,100. Joint distribution table:

$W_i \downarrow X_i \rightarrow$	0	1	$F(W_i)$
0	{3,5} 2/6 = 1/3	{4,6} 1/3	2/3
1	{1} 1/6	{2} 1/6	1/3
$P(X_i)$	1/2	1/2	1

One can see, for example, $P(W_i = 0|X_i = 1) = \frac{1}{3}$

$P(W_i = 0) = \frac{2}{3}$, $P(X_i = 1) = \frac{1}{2}$

Hence $P(W_i = 0)P(X_i = 1) = \frac{1}{3}$ for all i

An intervention to encourage physical activity is evaluated. Here are the weekly numbers of hours spent on exercising before and after the intervention.

	Person 1	Person 2	Person 3	Person 4	Person 5
Before	3	2.5	0	6	4
After	5	4	2	5	4.5

X: {3, 2.5, 0, 6, 4}, Y: {5, 4, 2, 5, 4.5},
$D = y - x$

a) *Test the hypothesis above with $\alpha = 0.05$. What is your conclusion?*
T test (d, alternative="greater"). One sample t-test. Data: d. t=1.7541, df=4, p-value=0.07714. Alternative hypothesis: true mean is greater than 0. 95% confidence interval: -0.2153397,∞. Sample estimates: mean of x 1. Do not Reject.

b) *Find a confidence interval with confidence level $1 - \alpha = 0.99$ for the expectancy of the weekly number of hours spent exercising after the intervention.* Answer: t.test(y, conf.level=0.99) -> 1.536547, 6.663452

c) *Test the hypothesis that the expectancy of weekly hours spent before the intervention is lower than 2.8. Use $\alpha = 0.05$. What is your conclusion?* Answer: t.test(x, alternative="less", $\mu = 2.8$). t=0.3062, df=4, p-value=0.6126. Do not Reject.

A publisher of an academic journal believes that the expectation of the number of pages of an academic paper has increases. To test his hypothesis, the publisher randomly samples six papers, that were published 20 years ago, and six that were published during the last year. Here are their length in number of pages.

Length (pages)	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5	Paper 6
Old Papers	22	38	31	29	32	20
New Papers	33	41	25	24	41	36

a) *Write the hypothesis of the publisher. What did you assume ? What is the p-value? What will you decide for $\alpha = 0.05$?*
Hypotheses: $H_0: \mu_{90s} = \mu_{last\ year}$. $H_1: \mu_{90s} < \mu_{last\ year}$. Test: Independent samples t-test. Assumptions: Each sample is identically and normally distributed. Each pair of observations is independent. Equal variance of both samples. $t(10) = -1.14$, $p - value = 0.1409 > 0.05 = \alpha$. Conclusion: Retain the null hypothesis.

b) *Eestimate the proportion of last-year academic papers with over 40 pages, using a confidence interval. What is the interval for 95% level of confidence?* Answer: PointEst: 0.33333, Lower: 0.09677141, Upper: 0.7000067

c) *A third publisher believes that the expectation of the length of papers in the 90s was lower than 35 pages. Write down the appropriate hypotheses and calculate the p-value. What will you decide for $\alpha = 0.10$. What did you assume?*
Hypotheses: $H_0: \mu_{90s} = 35$, $H_1: \mu_{90s} < 35$. Assumptions: Observations are identically and normally distributed. Each pair of observations is independent. $t(5) = -2.3212$, $p - value = 0.03 < 0.10 = \alpha$. Conclusion: Reject H_0.

The joint distribution of X and Y is described by the following joint density function:

$$f_{X,Y}(x,y) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), 0 < x < 1, 0 < y < 2 \\ 0, otherwise \end{cases}$$

a) *Find the marginal density, $f_X(x)$.* Answer: $f_X = \int_{-\infty}^{\infty} f_{X,Y}dy = \int_0^2 \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)dy = \cdots = \frac{3}{7}(4x^2 + x)$

b) *Calculate $P(X > Y)$. Answer: $P(X > Y); 0 \leq x \leq 1, -\leq y \leq x = \int_0^1\int_0^x \frac{6}{7}\left(x^2 + \frac{xy}{2}\right)dydx = \cdots = 0.268$*

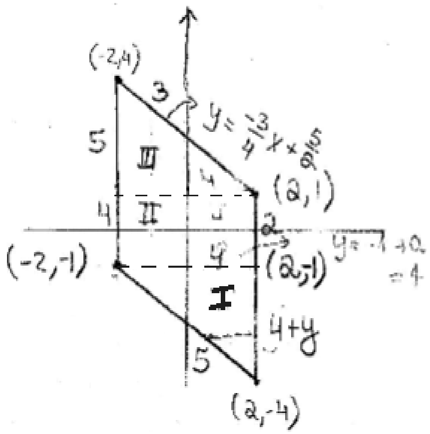
c) *Find the conditional distribution of Y given $X = x$.* Answer: $f_{Y|X} = \frac{f_{X,Y}}{f_X} = \frac{\frac{6}{7}x^2 + \frac{6}{14}xy}{\frac{12x^2}{7} + \frac{3}{7}x} = \cdots = \frac{2x^2 + xy}{4x^2 + x}$

d) *Are X and Y independent?* Answer: $f_Y = \frac{6}{7}\int_0^1\left(x^2 + \frac{xy}{2}\right)dx = \frac{6}{21} + \frac{y}{4}$, $f_X * f_Y = \left(\frac{12x^2}{7} + \frac{3}{7}x\right) * \left(\frac{6}{21} + \frac{y}{4}\right) \neq f_{X,Y} \rightarrow$ dependent.

A point, (X,Y) is chosen uniformly from the following parallelogram: (see augmented drawing below)

a) Does Y have a uniform distribution? Explain

Bottom edge=$y = -\frac{3}{4}x - \frac{5}{2}1 \cdot 4.$ -> $4y = -3x - 10$ -> $4y + 10 = -3x$, therefore $-\frac{4}{3}y - \frac{10}{3} = x$



The total area $A = 4 * 5 = 20 = \text{I} + \text{II}+\text{III}$

<u>I: $-4 \leq y \leq -1$</u>: (triangle) $S(\text{I}) = (2 - x)(4 + y)\frac{1}{2}$. $X = -\frac{4}{3}y - \frac{10}{3} \rightarrow$
$S(I) = \left(2 + \left(\frac{4}{3}y + \frac{10}{3}\right)(4 + y)\frac{1}{2} = \frac{16+4y}{3} \cdot (4 + y)\frac{1}{2} = \frac{64+16y+16y+4y^2}{6} \cdot \frac{1}{2} = \frac{32+16y+2y^2}{3}\right.$. $F_{\text{I}}(y) = S(\text{I})/20 = \frac{32+16y+2y^2}{60} = \frac{16+8y+y^2}{30} = \frac{(y+4)^2}{30}$, $-4 \leq y \leq -1$

<u>II: $-1 \leq y \leq 1$:</u> $S(\text{II}) = 4(y + 1)$. $S(\text{I}) + S(\text{II}) = \frac{4*3}{2} + (4(y + 1)$ $=10 + 4y$. Hence - $F(y) = \frac{10+4y}{20} = \frac{5+2y}{2}, -1 \leq y \leq 1$

<u>III: $1 \leq y \leq 4$:</u> $y = -\frac{3}{4}x + \frac{5}{2} \rightarrow 4y = -3x + 10 \rightarrow \frac{10-4y}{3} = x$. So
$S(\text{III}) = 6 - \left((4 - y) \cdot \left(\frac{10-4y}{2} - (-2)\right)\frac{1}{2}\right) = 6 - \frac{(4-y)}{2}\frac{10+4y+6}{3} = 6 - \frac{40+16y+24-10y+4y^2-6y}{6} = 6 - \frac{2y^2-16y+32}{3}$. Therefore, $S(\text{I}) + S(\text{II}) + S(\text{III}) = 6 + 8 + \left(6 - \frac{2y^2-16y+32}{3}\right)$. Hence, $F(y) = \frac{120 - \frac{2y^2-16y+32}{3}}{20}, 1 \leq y \leq 4$.

$$\text{Total: } F(y) = \begin{cases} \frac{(y+4)^2}{30}, -4 \leq y \leq -1 \\ \frac{5+2y}{2}, -1 \leq y \leq 1 \\ 1 - \frac{2y^2-16y+32}{60}, 1 \leq y \leq 4 \end{cases}$$

(Possible mistake in last eq-n. the "1" is probably "6")
NOT UNIFORM, as it is dependent on y

b) Find the marginal density of X

$f_X = \int_y \frac{1}{20}dy = \int_{-\frac{3}{4}x-2.5}^{-\frac{3}{4}x+2.5} \frac{1}{20}dy$ = as before $\frac{1}{20}\left[\frac{5}{2} * 2\right] = \frac{5}{20} = \frac{1}{4}$ uniform, $-\frac{4}{3}y + \frac{20}{6} \leq x \leq -\frac{4}{3} + \frac{20}{6}, -4 \leq y \leq 4$

c) *Find cond. density $f_{Y|X=x}$ for $x \in (-2,2)$.* Ans: $f_{Y|X=x} = \frac{f_{X,Y}(x,y)}{f_X(X=x)} = \frac{1/20}{1/4} = \frac{1}{5}, -4 \leq y \leq 4, -\frac{3}{4}x - 2.5 < y \leq -\frac{3}{4}x + 2.5$

d) What is the distribution of the random variable $E[Y|X]$. Answer: $(Y|X) \sim f_{Y|X}(y)$, $E(Y|X = x) = \int_y f_{Y|X} \cdot ydy = \int_{-\frac{3}{4}x-2.5}^{-\frac{3}{4}x+2.5} \frac{1}{5}ydy$ <- section c – uniform. $E(Y|X = x) = \frac{1}{5}\left[\frac{y^2}{2}\right]_{-\frac{3}{4}x-2.5}^{-\frac{3}{4}x+2.5} = -\frac{2.15}{10}x = -\frac{30}{40}x = \frac{3}{4}x, -2 \leq x \leq 2$ <- center of uniform. $E\left(E(Y|X)\right) = E\left(-\frac{3}{4}x\right) = -\frac{3}{4}E(X) = -\frac{3}{4}\int_x f_x x dx \rightarrow f_{E(Y|X)} = -\frac{3}{4}f_x = -\frac{3}{4}\frac{1}{4} = -\frac{3}{16}, -2 \leq x \leq 2, -\frac{3}{4}x - 2.5 < y \leq -\frac{3}{4}x + 2.5$.
$V(X) = \int x^2 f_x dx - (\int x f_x dx)^2 = E(X^2) - E^2(X) = \int_{-2}^2 \frac{1}{16}x^2 dx = \frac{1}{16}\left[\frac{x^3}{3}\right]_{-2}^2 = \frac{1}{48}(2^3 - (-2)^2) = \frac{16}{48} = 3$ (probable mistake here, should be 1/3, but keeping 3 for continuity) -> $= 3 - \left(-\frac{3}{16}\right)^2 = \frac{48-8}{16} = \frac{40}{16}$. Finally, $V\left(E(Y|X)\right) = \frac{90}{16}\frac{40}{16} = \frac{360}{256}$

10 Urns are given. Inside each urn, there are eight balls, number 1,2,…,8. From each urn one ball is randomly taken

a) *What is the probability that exactly three balls marked "3" are taken, and exactly two balls marked "4" are taken?* Answer:
$\frac{10!}{3!2!5!} \cdot \frac{1^5 \cdot 6^5}{8^{10}} = 0.01825$

b) *If it known that exactly three balls marked "1" are taken, what is the expectation of the number of balls marked "4" that are taken?*
Answer: Given the information, $Y \sim \text{Bin}\left(7, \frac{1}{7}\right)$, so the $E[Y] = 1$

c) *If it is known that exactly three balls marked "1" are taken, what is the probability that the third of those three balls was taken out of the eighth urn?* Answer: Let A be the event that a ball numbered "1" was taken out exactly three times, and B the event that the third of those three balls was taken out on the eighth urn. The wanted conditional probability is $P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{\binom{7}{2}\left(\frac{1}{8}\right)^3\left(\frac{7}{8}\right)^7}{\binom{10}{3}\left(\frac{1}{8}\right)^3\left(\frac{7}{8}\right)^7} = \frac{\binom{7}{2}}{\binom{10}{3}} = \frac{21}{120} = 0.175$

d) *What is the probability that "6" is the highest value mark taken out (it may appear on one or more balls)?* Answer: X –highest taken number. $P(X=6) = P(X\le 6) - P(X\le 5) = \left(\frac{6}{8}\right)^{10} - \left(\frac{5}{8}\right)^{10} = 0.04722$

The common distribution of X Y is described by the following distribution: $f_{X,Y}(x,y) = \begin{cases} xe^{-x(y+1)}, & 0<x,0<y \\ 0, & else \end{cases}$

a) *Calculate* $P\left(Y<\frac{1}{X}\right)$. Answer: $A = \left\{(x,y): x\ge 0, 0\le y\le \frac{1}{x}\right\}$. $P\left(Y\le\frac{1}{X}\right) = P\big((X,Y)\in A\big) = \iint_A f_{X,Y}(x,y)dxdy = \int_0^\infty xe^{-x}\int_0^{\frac{1}{x}} e^{-xy}dxdy = \int_0^\infty xe^{-x}\left[-\frac{1}{x}e^{-xy}\right]_{y=0}^{\frac{1}{x}} dx = \left(1-\frac{1}{e}\right)\int_0^\infty e^{-x}dx = 1-\frac{1}{e}$

b) *Calculate the variance of* X. Answer: $f_x(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = xe^{-x}\int_0^\infty e^{-xy}dy = e^{-x}$, for $x\ge 0$ $X\sim\text{Exp}(1)$, so $V(X)=1$

c) *Find the cond. distribution of Y|X=x*: Answer: $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{xe^{-x(y+1)}}{e^{-x}} = xe^{-xy}, xx>0, y>0$

d) *Calculate* $P((Y-3)^2>4)$. Answer: $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^\infty xe^{-x(y+1)}dx = = \left[-\frac{1}{1+y}xe^{-(y+1)x}\right]_{x=0}^{\infty} + \frac{1}{1+y}\int_0^\infty e^{-x(y+1)}dx = \frac{1}{(1+y)^2}, y\ge 0$. $F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt = \int_0^y \frac{dt}{(1+t)^2} = \frac{y}{1+y}, y\ge 0$. $P((Y-3)^2\ge 4) = P(Y-3>2) + P(Y-3<-2) = P(Y>5)+P(Y<1) = 1-F_Y(5)+F_Y(1) = 1-\frac{5}{6}+\frac{1}{2} = \frac{2}{3}$

e) *Are X and Y independent? Prove.* Answer: $f_X(x)\cdot f_Y(y) = e^{-x}\frac{1}{(1+y)^2} \ne f_{X,Y}(x,y)$, therefore dependent.

Ten balls numbered 1,…,10 are thrown into five cells numbered 1,…,5.

a) *What is the variance of the number of even numbered balls that ended in even numbered cells?* Answer: The variance of a binomial distribution with $n=5, p=0.4$ is $V(X) = 5*0.4*0.6 = 1.2$

b) *If it is known that three balls ended in cell number 1, what is the distribution of the number of even numbered balls that ended in cell number 1?* Answer: $Y\sim\text{HG}(N=10, n=3, D=5)$

c) *A ball is called "ordered" if in the cell that is one number below its, there is a ball with the number that is one below its own number. Thus, ball number 7 will be called "ordered" if it ends, for example, in cell5, and ball number 6 ends in cell 6. Find the expectation and variance of the number of "ordered" balls.* Answer: We give an indicator for any ball numbered 2 to 10. The probability of being "ordered" is $\frac{4}{26}$ (choose 2 cells, $\left(\frac{1}{5}\right)*4$ pairs). SO the expectation of the number of ordered balls is $\frac{9*4}{25} = \frac{36}{25}$ (possible error, denominator changed from 26 to 25). In order to find the variance, we need to find the covariance of every pair of indicators. If the numbers of the balls are not consecutive, then the indicators are independent. If the pair of balls has consecutive numbers, then $P(I_i=1, I_{i+1}=1) = \frac{3}{5^3}$. Therere are 8 such pairs with the covariance $\frac{3}{5^3} - \frac{16}{5^4} = -\frac{1}{5^4}$. Thereofre, the variance of sum of the indicators is $n(pairs)*p*(1-p) - 2\text{Cov}(X,Y) = 9*\frac{4}{25}*\frac{21}{25} - 2*8*\frac{1}{5^4} = 1*\frac{23}{125} = 1.184$

The teacher Shmuel wrote two bible exams, exam A for class 1, and exam B for class 2. Here is a sample of 5 grades from each class: class 1: 85,74,64,42,30. Class 2: 71,70,56,38,25

a) *The students in class 2 claim their exam was harder and the grades expectation is lower than class 1. Test that claim with* $\alpha = 0.05$. *What do you assume?* Answer: Assume equal variance and normal distribution. Perform t-test for independent samples. For class 1: $\bar{X_1} = 58.4$, $s_! = 22.5$. For class 2: $\bar{X_2} = 52, s_2 = 20.16$. $S^2_{pooled} = \frac{(n_1s_1^2)+(n_2s_2^2)}{n_1+n_2-2} = 456.33$. $T = \frac{\overline{X_2}-\overline{X_1}}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{58.4-52}{\sqrt{456.33*0.4}} = 0.473$.

Looking at t-table, the row of 8df, $t_{8,0.95}$=1.86, so we can't reject the null hypothesis .

b) *The students of class 1 claim that the proportion of failures (less than 60 is a fail) is significantly higher than 10%. Test their claim with significance level 5%. Can we justify them?* Answer; The failure proportion of class 1 is 0.4. Perform Z-test: $Z = \frac{p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.4-0.1}{\sqrt{\frac{0.1*0.9}{5}}} = 2.23$. Looking at z-table, we find that $Z_{0.95} = 1.645$. We can reject the null hypothesis, indeed the failing proportion is significantly higher than 10%

c) *Calculate a confidence interval with confidence level of 95%, for the expectation of grades in class 1. What did you assume?* Answer: $\bar{X} \pm t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}} = 58.4 \pm 2.776\ \frac{22.5}{\sqrt{5}}$

Selling products. Profit b for each sold, loss of l for each unit left unsold. Number of units sold within. A season is a random variable X with probability mass function P(X=i)=P(i). determine number of products store should stock to maximize profits. Number of units stocked:s total profit: $\pi(s)\begin{cases} bX-(s-X)l; & X\le s \\ bs; & X>s \end{cases}$ increasing stock is profitable when: $\sum_{i=0}^{s} P(X=i) < \frac{b}{b+l}$

Misc. Formulae:

$F_y(y) = \int_{-\infty}^{y} f_y(t)dt$	$P(Y>5)+P(Y<1) = 1-F_y(5)+F_y(1)$
$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$	$E\big[E[X\|Y]\big] = \sum(\text{Total } P(y)s) * (\text{weighted average of } x)$
$E[XY] = \sum P(XY=xy)*xy$	$P\left(-\frac{1}{3}<Z<\frac{2}{3}\right) = \Phi\left(\frac{2}{3}\right)-\Phi\left(-\frac{1}{3}\right) = \Phi\left(\frac{2}{3}\right)-\left(1-\Phi\left(\frac{1}{3}\right)\right)$