University of California, Berkeley STAT 134: Concepts of Probability Summary

Chapter 0: Prerequisite	
Set Theory	**Definition**: A set if an unordered collection of elements. It looks like $S = \{X \in A: "property"\}$. **Some basic properties and identities about sets:** Basics: 1. $\boxed{\emptyset}$ means an empty set. 2. $\boxed{a \in A}$ means the object a is an element of the set A. 3. $\boxed{A \subseteq B}$ means A is a subset of B. 4. $\boxed{A = B \Leftrightarrow A \subseteq B\ and\ B \subseteq A}$. 4. $\boxed{\emptyset \subseteq A}$ for every set A. Operations: 1. $\boxed{A \cup B}$ is called the union of A and B, whose elements are the elements of A and B. 2. $\boxed{A \cap B}$ is called the intersection of A and B, whose elements are the elements common to A and B. 3. $\boxed{A - B}$ is called the difference of A and B, whose elements are those elements of A that are not members of B. 4. Associativity: $\boxed{A \cup (B \cup C) = (A \cup B) \cup C}$, $\boxed{A \cap (B \cap C) = (A \cap B) \cap C}$. 5. Commutativity: $\boxed{A \cup B = B \cup A}$, $\boxed{A \cap B = B \cap A}$. 6. Distributivity: $\boxed{A \cup (B \cap C) = (A \cup B) \cap (A \cup C)}$, $\boxed{A \cap (B \cup C) = (A \cap B) \cup (A \cap C)}$. 7. Idempotency: $\boxed{A \cup A = A}$, $\boxed{A \cap A = A}$, $\boxed{A - A = \emptyset}$. 8. Identity: $\boxed{A \cap \Omega = A}$, $\boxed{A \cup \emptyset = A}$. 9. Domination: $\boxed{A \cup \Omega = \Omega}$, $\boxed{A \cap \emptyset = \emptyset}$. 10. Complementation: $\boxed{(A^c)^c = A}$, $\boxed{A = 1 - A^c}$, $\boxed{A \cup A^c = \Omega}$, $\boxed{A \cap A^c = \emptyset}$. 11. De Morgan's: $\boxed{(A \cap B)^c = A^c \cup B^c}$, $\boxed{(A^c \cup B^c) = A^c \cap B^c}$.

	12. Absorption: $\boxed{A \cup (A \cap B) = A}$, $\boxed{A \cap (A \cup B) = A}$. Notations: 1. Union: $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \dots \cup A_{n-1} \cup A_n$. 2. Intersection:$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n$.
Counting	**1. One-to-one Correspondence** If A is a set consists of finite elements, there is a one-to-one correspondence between the elements in A and elements B. Then we know that B is finite and A and B have the same number of elements. **2. Addition Principle** If a first task can be done in n_1 ways and a second task can be done in n_2 ways, and if these tasks cannot be done at the same time, then there are $\boxed{n_1 + n_2}$ ways to do both tasks. **3. Multiplication Principle** If experiment A has m outcomes and B has n outcomes, then the sequence of experiments (A, B) has $\boxed{m \times n}$ outcomes. **4. Permutations** A permutation is an **ordered** arrangement of objects. It is well-known as the number of orderings for elements. If we have a set $S = \{A_1, A_2, \dots, A_{n-1}, A_n\}$, and we need to choose r elements and put them in order, then the number of orderings is as follow. 1) Without Replacement $\boxed{P(n,r) = \frac{n!}{(n-r)!} = n \times (n-1) \times \dots \times (n-r+1)}$ If we do permutation with k repeated elements, each element contains $a, b, c, \dots, \infty$ items repeated, we use the following formula $\boxed{\frac{P(n,r)}{a!\cdot b!\cdot c!\cdot \dots} = \frac{\frac{n!}{(n-r)!}}{a!\cdot b!\cdot c!\cdot \dots} = \frac{n\times(n-1)\times \dots \times(n-r+1)}{a!\cdot b!\cdot c!\cdot \dots}}$ 2) With Replacement $\boxed{P^R(n,r) = n^r}$ **5. Combinations** A combination is an **unordered** arrangement of objects. It is well-known as the binomial coefficient. If we have a set $S = \{A_1, A_2, \dots, A_{n-1}, A_n\}$, and we need to choose r elements, no matter what the order is, then the number of combinations is as follow. 1) Without Replacement $\boxed{C(n,r) = \binom{n}{r} = \frac{P(n,r)}{P(r,r)} = \frac{n!}{r!(n-r)!}}$ 2) With Replacement $\boxed{C^R(n,r) = \frac{(n+r-1)!}{r!(n-r)!}}$

Chapter 1: Introduction to Probability	
Definition	To define probability, we start with an outcome space Ω, and assign to each element a nonnegative number and require that all numbers add up to 1.
Axioms	We define Ω as the outcome space and A as one of its events. **Axiom I:** $P(A) \geq 0$ for all $A \subseteq \Omega$. **Axiom II:** $P(\Omega) = 1$. **Axiom III:** If A and B have no element in common, then $P(A \cup B) = P(A) + P(B)$. A system of outcome spaces together with a defined assignment of numbers $P(A)$, satisfying all the axioms above, is called a field of probability.
Equally Likely Outcomes	If all outcomes in a finite set Ω are equally likely, the probability of A is the number of outcomes in A divided by the total number of outcomes, which is $\boxed{P(A) = \frac{number\ of\ A}{number\ of\ \Omega}}$. Probability defined by this formula for equally likely outcomes are fractions between 0 and 1. The number of 1 represents certainty, which means $\boxed{P(\Omega) = 1}$. The number of 0 represents impossibility, which means $\boxed{P(\emptyset) = 0}$.
Interpretations	**1. Frequency** Long run proportion of occurrence converges to a number, and this number is probability. $P_n(A) = P(A)$ for large n. **2. Baysian/Subjective** Probability if a degree of certainty of an outcome. **3. Mathematics** Probability is a function that takes in collections of outcomes and give a number. **Difference between Statistics and Probability:** In general, Statistics and Probability are inverse. Statistics is that given the sample we have, we want to predict the population. However, probability is that given the population, we want to predict the sample we will get.
Basic Probability Rules	1. $\boxed{0 \leq P(A) \leq 1}$. 2. $\boxed{P(\Omega) = 1}$. 3. $\boxed{P(\emptyset) = 0}$. 4. $\boxed{P(A \cap B) = P(A\ and\ B) = P(both\ event\ A\ occurs\ and\ event\ B\ occurs)}$. 5. $\boxed{P(A \cup B) = P(A\ or\ B) = P(event\ A\ occurs\ or\ event\ B\ occurs\ or\ both\ occur)}$. 6. $\boxed{P(A \cap B) = 0}$ if event $\boxed{A \text{ and event } B \text{ have no element in common}}$. 7. $\boxed{P(A \cup A^c) = P(\Omega) = 1}$. 8. $\boxed{P(A \cap A^c) = P(\emptyset) = 0}$.

Addition Rule	Suppose we have an outcome space Ω, and we also have two events A and B. Assume A and B have no element in common. $\boxed{P(A \cup B) = P(A) + P(B)}$ Proof: We know that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (inclusion-exclusion rule). We also know that $P(A \cap B) = P(\emptyset) = 0$ when A and B have no element in common (basic rule of probability). As a result, $P(A \cup B) = P(A) + P(B)$. ■
Difference Rule	Suppose we have an outcome space Ω, and we also have two events A and B. Assume $B \subseteq A$. $\boxed{P(A \backslash B) = P(A) - P(B)}$ Proof: Since $B \subseteq A$, then we know that $P(A \cap B) = P(B)$, $P(A \cup B) = P(A)$. We also know that $P(A \backslash B) = P(A \cap B^c)$ (basic probability conversion). $P(A) = P(A \cap B) + P(A \cap B^c) = P(A) + P(B) - P(A \cap B) = P(A \cup B) \Rightarrow P(B) + P(A \cap B^c) = P(A) \Rightarrow P(A \cap B^c) = P(A) - P(B)$. As a result, $P(A \backslash B) = P(A) - P(B)$. ■
Partitioning Rule	Suppose we have an outcome space Ω, and we also have two events A and B. (It is easy to prove this by using Ven Diagram) $\boxed{P(A) = P(A \cap B) + P(A \cap B^c)}$
Complement Rule	Suppose we have an outcome space Ω, and we also have one event A. $\boxed{P(A^c) = 1 - P(A)}$ Proof: We know that $P(A \cup A^c) = P(A) + P(A^c) - P(A \cap A^c) = P(\Omega) = 1$ (inclusion-exclusion rule and Axiom II). We also know that $P(A \cap A^c) = P(\emptyset) = 0$ (basic rule of probability). As a result, $P(A \cup A^c) = P(A) + P(A^c) = 1 \Rightarrow P(A) = 1 - P(A^c)$. ■
Inclusion-Exclusion Rule	Suppose we have an outcome space Ω, and we also have two events A and B, which have elements in common. $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$ Proof: We can consider $P(A \cup B) = P(A) + P(B \cap A^c)$ (basic probability conversion). We can also know that $P(B) = P(B \cap A) + P(B \cap A^c) \Rightarrow P(B \cap A^c) = P(B) - P(B \cap A)$ (partitioning). As a result, $P(A \cup B) = P(A) + P(B) - P(B \cap A)$. ■ **General Formula of Inclusion-Exclusion:** $\boxed{P(\cup_{i=1}^{n} A_i) = P(\Sigma_{i=1}^{n} A_i) - P\left(\Sigma_{i<j\leq n} A_i \cap A_j\right) + P\left(\Sigma_{i<j<k\leq n} A_i \cap A_j \cap A_k\right) + \cdots + (-1)^{n+1}(A_i \cap A_j \cap \ldots \cap A_n)}$ Proof: **Base cases:** When $n = 1$, $P(\cup_{i=1}^{1} A_i) = P(A_1)$. When $n = 2$, $P(\cup_{i=1}^{2} A_i) = P(\Sigma_{i=1}^{2} A_i) - P(\Sigma_{i<j\leq 2} A_i \cap A_j) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. **Inductive hypothesis:** We assume that $P(\cup_{i=1}^{n} A_i) = P(\Sigma_{i=1}^{n} A_i) - P\left(\Sigma_{i<j\leq n} A_i \cap A_j\right) + P\left(\Sigma_{i<j<k\leq n} A_i \cap A_j \cap A_k\right) + \cdots + (-1)^{n+1}(A_i \cap A_j \cap \ldots \cap A_n)$ is always true for positive integer n. **Inductive step:** $P\left(\cup_{i=1}^{n+1} A_i\right) = P(\cup_{i=1}^{n} A_i \cup A_{n+1}) = P(\cup_{i=1}^{n} A_i) + P(A_{n+1}) - P(\cup_{i=1}^{n} A_i \cap A_{n+1})$ (use the base case). As a result, the claim is true for all n. ■

Conditioning	**1.** We usually denote the following probability as "**The probability of A given B**." $\boxed{P(A\mid B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B\cap A)}{P(B)}}$ **2.** By the multiplication rule, we have the following formula. $\boxed{P(A\cap B)=P(B\cap A)=P(B)\cdot P(A\mid B)=P(A)\cdot P(B\mid A)}$ **3.** By partitioning, we have the following formula. $\boxed{P(A)=P(A\cap B)+P(A\cap B^c)=P(B)\cdot P(A\mid B)+P(B^c)\cdot P(A\mid B^c)}$ **4.** Suppose we have n events, which are $A_1,A_2,\dots,A_n$, then we have the multiple events conditioning multiplication rule as following. $\boxed{P(A_1\cap A_2\cap\dots\cap A_n)=P(A_1)\cdot P(A_2\mid A_1)\cdot P(A_3\mid A_1\cap A_2)\cdot\dots\cdot P(A_n\mid A_1\cap A_2\cap\dots\cap A_{n-1})}$
Independence	We say A and B are **independent** in any of the following cases. **1.** $\boxed{P(A\mid B)=P(A)}$ **2.** $\boxed{P(B\mid A)=P(B)}$ **3.** $\boxed{P(A\mid B)=P(A\mid B^c)}$ **4.** $\boxed{P(B\mid A)=P(B\mid A^c)}$ **5.** $\boxed{P(A\cap B)=P(A)\cdot P(B)}$ **Independence vs Mutually Exclusive** Mutually exclusive: $P(A\cap B)=0$. Independent: $P(A\cap B)=P(A)\cdot P(B)$. Both mutually exclusive and independent: $\boxed{P(A\cap B)=P(A)\cdot P(B)=0}$ (at least one of $P(A)$ and $P(B)$ is zero).
Pairwise, Mutually, and Three-events Independence	We say events A,B,C are <u>pairwise independent</u> **if and only if** $P(A\cap B)=P(A)\cdot P(B)$, $P(A\cap C)=P(A)\cdot P(C)$, and $P(B\cap C)=P(B)\cdot P(C)$. We say events A,B,C are <u>mutually independent</u> **if and only if** $P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)$. As a result, events A,B,C are **<u>independent</u> if and only if** they are **<u>(1) pairwise independent</u>** ($P(A\cap B)=P(A)\cdot P(B)$, $P(A\cap C)=P(A)\cdot P(C)$, and $P(B\cap C)=P(B)\cdot P(C)$) and **<u>(2) mutually independent</u>** ($P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)$).
Bayes' Rule	$\boxed{P(A\mid B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B\mid A)\cdot P(A)}{P(B\mid A)\cdot P(A)+P(B\mid A^c)\cdot P(A^c)}}$ In general, let $A_i=\{A_1,A_2,\dots,A_n\}$ and B be events, where A_i are disjoint, $\cup_{i=1}^{n}\ A_i=\Omega$, and $P(A_i)>0$ for all i. Then we get the general formula $\boxed{P(A_j\mid B)=\frac{P(B\mid A_j)\cdot P(A_j)}{\Sigma_{i=1}^{n}P(B\mid A_i)\cdot P(A_i)}}$.

Chapter 2: Repeated Trials, Sampling, and Distributions	
Uniform Distribution (continuous)	**1. Uniform distribution on a finite set** This is the distribution of a point picked at random from a finite set Ω, so that all points are equally likely to be picked. For $A \subset \Omega$. $\boxed{P(A) = \frac{\#\,(A)}{\#\,(\Omega)}}$. **Special cases:** **1.** Bernoulli (1/2) distribution: uniform distribution on {0,1}. **2.** The number of a fair-die roll: uniform distribution on {1,2,3,4,5,6}. **3.** Uniform distribution on $\{1,2,\ldots,n\}$: Let X have uniform distribution on the integer 1 to n. Then $E(X) = \frac{n+1}{2}$ and $SD(X) = \sqrt{\frac{n^2-1}{12}}$. 2. Uniform distribution on an interval We define $a < b$ as the endpoints of the interval, which of course the range is $x \in (a, b)$. **Density function:** $\boxed{\frac{P(X \in dx)}{dx} = \frac{1}{b-a}}$, where $a \leq x \leq b$. The probability of any subinterval of (a, b) is proportional to the length of the subinterval. **Cumulative distribution function:** $\boxed{P(X \leq x) = 0, if\ x < a}$ $\boxed{P(X \leq x) = \frac{x-a}{b-a}, if\ a \leq\ x \leq b}$ $\boxed{P(X \leq x) = 1, if\ x > b}$ **Expectation/Mean:** $\boxed{E(X) = \frac{a+b}{2}}$ **Standard deviation:** $\boxed{SD(X) = \sqrt{\frac{(b-a)^2}{12}}}$
Bernoulli Distribution (discrete, independent, 2 outcomes)	Bernoulli distribution is used when there are only **two outcomes**, success (p) and failure (1−p). We denote Bernoulli distribution as $\boldsymbol{Ber(p)}$. $\boxed{P(1) = p, P(0) = 1 - p}$ Expectation/Mean: p Standard deviation: $np(1 - p)$

Binomial Distribution <u>(discrete, independent, 2 outcomes)</u>

Suppose we have n **independent (with replacement)** $Ber(p)$ trials ($n = 1,2, \ldots$), with p is the probability of success on each trial ($0 \leq p \leq 1$). We define k as the range $\{0,1,2, \ldots, n\}$. We denote Binomial distribution as $\boldsymbol{Bin(n,p)}$.

Probability function:

$\boxed{P(k) = P(S = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}}$ is the chance of have each number (k) of success, where $S =$ (number of successes in n independent trials with probability p of success on each trial) $= X_1 + X_2 + \cdots + X_n$, where $X_i =$ indicator of success on trial i.

The total chance of having all the number (from 1 to k) of success is $\boxed{\Sigma_{k=0}^{n} P(k) = \Sigma_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = 1}$. This is also how we calculate the exact probability from a binomial distribution.

Mode: $\boxed{Mode\ of\ Bin(n,p) = \lfloor np + p \rfloor}$

1) $k < np + p$ if and only if $P(k-1) < P(k)$.
2) $k = np + p$ if and only if $P(k-1) = P(k)$.
3) $k > np + p$ if and only if $P(k-1) > P(k)$.

Expectation/Mean: $\boxed{Mean\ of\ Bin(n,p) = E(S) = \mu = np}$

Standard deviation: $\boxed{SD(S) = \sqrt{Var(X)} = \sqrt{\sigma^2} = \sigma = \sqrt{np(1-p)}}$

Consecutive odd ratio: $\boxed{\frac{P(k)}{P(k-1)} = \frac{\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}}{\binom{n}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-k+1}} = \frac{n-k}{k+1} \cdot \frac{p}{1-p}}$ (decreasing)

Special case: $Bin(1,p) \equiv Ber(p)$, distribution of the indicator of an event A with probability $P(A) = p$.

<u>Approximations:</u>

1. Normal Approximation

It is difficult to calculate the exact probabilities with the binomial format. It is easier to calculate the area under the normal curve. $Bin(n,p) \approx N(np, \sqrt{np(1-p)})$ when $n \geq 20$ and $\sigma = \sqrt{np(1-p)}$ is sufficiently large.

Then $\boxed{P(k) \approx \frac{1}{\sigma} \cdot \phi\left(\frac{k-\mu}{\sigma}\right) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot \left(\frac{k-\mu}{\sigma}\right)^2}}$, where $\phi\left(\frac{k-\mu}{\sigma}\right) = \phi(z)$ is the standard normal density function.

We also have $\boxed{P(a\ to\ b) \approx \Phi\left(\frac{b+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{a-\frac{1}{2}-\mu}{\sigma}\right)}$, where Φ is the standard normal cumulative distribution function.

# Notice here we use continuity correction (± 0.5) since binomial distribution is discrete and we are using normal distribution, which is continuous, to approximate it.

	We can get a confidence interval using the normal approximation $\mu \pm z_{1-\frac{\alpha}{2}}\sigma = np \pm z_{1-\frac{1}{2}}\sqrt{np(1-p)} = \boxed{p \pm \frac{z_{1-\frac{1}{2}}\sqrt{p(1-p)}}{\sqrt{n}}}$ or $\boxed{\hat{p} \pm \frac{z_{1-\frac{1}{2}}\sqrt{p(1-p)}}{\sqrt{n}}}$. # If n is large, it is natural to expect that the unknown probability p is most likely fairly close to $\hat{p}$. **2. Poisson Approximation** For cases when p is very small (or p is close to 1) and n is large, we can use Poisson distribution to approximate Binomial distribution. $Bin(n, p) \approx Pois(no) = Pois(\mu)$. We have the chance $\boxed{P(k) \approx e^{-\mu} \cdot \frac{\mu^k}{k!}}$ for each number (k) of success. # Notice here we don't use continuity correction (± 0.5) since both Binomial distribution and Poisson distribution are discrete.
Multinomial Distribution (discrete, independent, k outcomes)	Suppose we have n **independent (with replacement)** trials ($n = 1,2, \ldots$). We define k as the range $\{0,1,2,\ldots,n\}$ and $n_i = \{n_1, n_2, \ldots, n_k\}$. For each trial of n_i, we have p is the probability of success ($0 \le p \le 1$). $\boxed{P(n_1, n_2, \ldots, n_k) = \binom{n}{n_1, n_2, \ldots, n_k} \cdot p^{n_1} \cdot p^{n_2} \cdot \ldots \cdot p^{n_k}}$
Hypergeometric Distribution (discrete, dependent, 2 outcomes)	Suppose we have n **dependent (without replacement)** trials ($n = 1,2, \ldots$). We define N as the total population size, G as the number of good elements in the population, g as the number of good elements in the sample n. $N = G + B$ and $n = g + b$. We denote Hypergeometric distribution as $\boldsymbol{HG(N, G, n)}$. Let S be the distribution of the number. **Probability function:** $\boxed{P(g) = P(S = g) = \binom{n}{g}\frac{(G)_g(B)_b}{(N)_n} = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}}$ is the chance of getting g good elements and b bad elements in the random sample of size n. The random variable is $S =$ number of good elements in sample $= X_1 + \cdots + X_n$, where X_i = indicator of the event that the $i$thh element sampled is good. These indicators are dependent due to sampling without replacement. But each indicator has the same $Ber(p)$ distribution, where $p = \frac{G}{N} = P(X_i = 1) = P(i\text{th element is good})$ for each $i = 1, \ldots, n$. **Expectation/Mean:** $\boxed{Mean\ of\ HG(H, G, n) = E(S) = np = \mu}$ **Standard deviation:** $\boxed{SD(S) = \sqrt{Var(S)} = \sqrt{\sigma^2} = \sigma = \sqrt{np(1-p)} \cdot \sqrt{\frac{N-n}{N-1}}}$ # Mean is the same as for sampling with replacement. But the SD is decreased by the correction factor of $\sqrt{\frac{N-n}{N-1}}$.

Multivariate Distribution (discrete, dependent, k outcomes)	Suppose we have n **dependent (without replacement)** trials ($n = 1,2,...$). We define N as the total population size, $G = \{G_1, ..., G_k\}$ as the number of good elements in the population, $g_i = \{g_1, ..., g_k\}$ as the number of good elements in the sample n. $N = G_1 + \cdots + G_k$ and $g = g_1 + \cdots g_2$. **Probability function:** $\boxed{P(g_1, g_2, ..., g_k) = \frac{\binom{G_1}{g_1}\cdot\binom{G_2}{g_2}\cdot\ldots\cdot\binom{G_k}{g_k}}{\binom{N}{n}}}$
Normal Distribution (continuous)	**1. Standard Normal: $N(0,1)$** **Standard normal density function:** $\boxed{\frac{P(Z\in dz)}{dz} = \phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}}$ for $z \in (-\infty, \infty)$. **Standard normal cumulative distribution function (CDF):** $\boxed{P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \phi(x)dx}$. By transformation, we get $\boxed{\Phi(-z) = 1 - \Phi(z)}$. **Expectation/Mean:** 0 **Standard deviation:** 1 **2. Normal: $N(\mu, \sigma^2)$** We denote $X = \mu + Z\sigma$, where $Z \sim N(0,1)$. All formulae follow from this linear change of variable. **Density function:** $\boxed{\frac{P(X\in dx)}{dx} = \frac{1}{\sigma}\cdot\phi\left(\frac{x-\mu}{\sigma}\right)}$ for $x \in (-\infty, \infty)$. **Cumulative distribution function (CDF):** $\boxed{P(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$ for $x \in (-\infty, \infty)$. **Expectation/Mean:** μ **Standard deviation:** σ
Poisson Distribution (discrete)	We define μ as the mean number and k is the range $\{0,1,2,...\}$. We denote Poisson distribution as $\boldsymbol{Pois(\mu)}$. **Probability function:** $\boxed{P(k) = P(N_\mu = k) = e^{\mu}\cdot\frac{\mu^k}{k!}}$ for $k = 0,1,2,....$ N_μ is the number of arrivals in a given time period in a Poisson arrival process, of the number of points in a given area in a Poisson random scatter, when the expected number is μ. **Expectation/Mean:** $\boxed{Mean\ of\ Pois(\mu) = E(N_\mu) = \mu}$ **Standard deviation:** $\boxed{SD(\mu) = \sqrt{Var(S)} = \sqrt{\mu}}$

Geometric Distribution (discrete)

We define p as the success probability with the range $n \in \{1,2,\ldots\}$. It is the distribution of the waiting time T to first success in independent trials with probability p success on each trial.

Probability function: $\boxed{P(n) = P(T = n) = (1-p)^{n-1}p}$, where $n \in \{1,2,\ldots\}$. Let $F = T - 1$ denote the number of failures before the first success. The distribution of F is the geometric distribution on $\{0,1,2,\ldots\}$.

Tail probabilities: $\boxed{P(T > n) = P(\text{first } n \text{ trials are failures}) = (1-p)^n}$

Expectation/Mean: $\boxed{Mean = E(T) = \frac{1}{p}}$ **Standard deviation:** $\boxed{SD(T) = \sqrt{Var(T)} = \frac{\sqrt{1-p}}{p}}$

Negative Binomial (discrete)

We define p as the success probability with the range $n \in \{0,1,2,\ldots\}$. Moreover, we define r as the number of successes. It is the distribution of the number of failures F_r before the rth success in Bernoulli trials with probability p of success on each trial.

Probability function: $\boxed{P(F_r = n) = P(T_r = n + r) = \binom{n+r-1}{r-1}p^r(1-p)^n}$, where $n \in \{0,1,2,\ldots\}$, and T_r is the waiting time to the rth success. The distribution of $T_r = F_r + r$ is the negative binomial distribution on $\{r, r+1,\ldots\}$.

Expectation/Mean: $\boxed{Mean = E(F_r) = \frac{r(1-p)}{p}}$ **Standard deviation:** $\boxed{SD(F_r) = \sqrt{Var(F_r)} = \frac{\sqrt{r(1-p)}}{p}}$

Sum of geometrics: The sum of r independent geometric (p) random variables on $\{0,1,2,\ldots\}$ has negative binomial (r,p) distribution.

Chapter 3: Random Variables

Random Variable

A random variable, X, is the outcome of an experiment, from outcomes to real numbers. For example, let's flip a two-sided coin two times. $X =$ the number of heads. $X = 1$ is an event that there is only one head. $P(X = 1) = \binom{2}{1} \cdot p^1 \cdot (1-p)^1$. We can write is as $Bin(2,p)$.

$X\colon\ \Omega \Rightarrow R$ (This is a function)
$HH \Rightarrow 2$
$HT \Rightarrow 1$
$TH \Rightarrow 1$
$TT \Rightarrow 0$

X has a probability **distribution**:

x	0	1	2
$P(X)$	1/4	1/2	1/4

Joint Distribution

Let (X, Y) be the joint outcome of 2 random variables X, Y. The event $(X = x, Y = y)$ is the intersection of events $X = x$ and $Y = y$. For example, we toss a coin twice and we have a random variable X = number of heads and a random variable Y = number of tails. We know that $\boxed{P(X = x \cap Y = y) = P(Y = y|X = x) \cdot P(X = x)}$.

$\boldsymbol{Y \setminus X}$	**0**	**1**	**2**	**Dist. Y**
0	0	0	$1/4$	$1/4 + 0 + 0 = \mathbf{1/4}$
1	0	$1/2$	0	$0 + 1/2 + 0 = \mathbf{1/2}$
2	$1/4$	0	0	$0 + 0 + 1/4 = \mathbf{1/4}$
Dist. X	$0 + 0 + 1/4 = \mathbf{1/4}$	$0 + 1/2 + 0 = \mathbf{1/2}$	$1/4 + 0 + 0 = \mathbf{1/4}$	

Distribution of Y

Marginals of Y

Distribution of X Marginals of X

The joint table of X and Y also means the distribution of (X, Y). The marginal probability of Y is also the distribution of Y, which is $\boxed{P(Y) = \Sigma_{x \in X} P(X, Y)}$. The marginal probability of X is also the distribution of X, which is $\boxed{P(X) = \Sigma_{y \in Y} P(X, Y)}$.

If the random variables X, Y are **independent**, the **diagonal** of the table will be **nonnegative** numbers and others are zero.

Properties of Random Variable

Suppose f and g are two functions, $f = g$ **if and only if** $f(x) = g(x)$ for all x.

If random variables $X = Y$, then they have the same distribution $\boxed{P(X = \omega) = P(Y = \omega)}$.

If X has the same distribution of Y, then any statement about X has the same probability as the corresponding statement about Y, and $g(X)$ has the same distribution as $g(Y)$, for any function g. $\boxed{P(a \le X \le b) = P(a \le Y \le b)}$ for all a and b. This also holds for any other forms such as $\boxed{P(a \le X^2 \le b) = P(a \le Y^2 \le b)}$.

For each possible value x of X, as y varies over the range of Y, the probability $P(Y = y|X = x)$ defines a probability distribution over the range of Y. This probability distribution, which may depend on given value of x of X, is called the conditional distribution Y given $X = x$. By the multiplication rule, we can get $\boxed{P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y|X = x)}$.

If X and Y are independent, then $\boxed{P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)}$ for all x and y.

The following three conditions are equivalent:
1. X and Y are independent.
2. The conditional distribution of Y given $X = x$ does not depend on x.
3. The conditional distribution of X given $Y = y$ does not depend on Y.

The distribution of X is **symmetric about 0** if $\boxed{P(X = -x) = P(X = x)}$ for all x.

Expectation	**Definition:** Expectation is an operator that takes a random variable and returns its mean, which is a real number. X must take values in real numbers. It is also called mean of first moment. If X takes finitely many values, then $\boxed{E(X) = \Sigma_k k \cdot P(X = k)}$, where k ranges all values taken by x. As a result, $\boxed{\min(x) \leq E(X) \leq \max(x)}$. **Linearity of expectation:** **1.** $\boxed{E(c) = c}$, where c is a constant. **2.** $\boxed{E(X + Y) = E(X) + E(Y)}$. # X, Y do not need to be independent. **3.** $\boxed{E(aX + b) = aE(X) + b}$, where a and b are constants. If random variables X and Y have the same distribution, then $\boxed{E(X) = E(Y)}$. If random variables X and Y are independent, then $\boxed{E(XY) = E(X) \cdot E(Y)}$. **Expectation of a function of a random variable:** $E(X) = \Sigma_{x \in X} x \cdot P(X = x) \Rightarrow \boxed{E\big(g(X)\big) = \Sigma_{x \in X} g(x) \cdot P(X = x)}$ **Expectation of (X, Y) joint distribution:** $\boxed{E\big(g(X, Y)\big) = \Sigma_{all\ x,y}\, g(x, y) \cdot P(X = x, Y = y)}$ **Indicator:** An indicator is **a random variable** that has **only two values 1 (with probability p) and 0 (with probability $1 - p$)**. Then $E(I) = 1 \cdot p + 0 \cdot (1 - p)$. Random variables that are counts can often written as a sum of indicators.
Markov Inequality	Let $X \geq 0$, $a \geq 0$, $E(X)$ is the center of the distribution, we then have our **Markov inequality** $\boxed{P(X \geq a) \leq \frac{E(X)}{a}}$. **Tail sum:** Suppose X is a random variable and takes values 0, 1, 2, …, n. Then we have $E(X) = \Sigma_{k=0}^{n} k \cdot P(X = k) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + \cdots + n \cdot P(X = n)$. Since $0 \cdot P(X = 0) = 0$. As a result, we have $\boxed{E(X) = \Sigma_{k=1}^{n} P(X \geq k) = P(X \geq 1) + P(X \geq 2) + \cdots + P(X \geq n)}$.
Variance and Standard Deviation	**Definition:** Variance is a measure of fluctuation. Standard deviation is the average spread of the data around the mean. Variance of a random variable is also called the **second central moment** of a random variable $\boxed{Var(X)} = E(X^2 - \mu) = E(X^2 - 2X\mu + \mu^2) = E(X^2) - E(2X\mu) + \mu^2 = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = \boxed{E(X^2) - \big(E(X)\big)^2}$ **OR** $\boxed{Var(X) = E\left(\big(X - E(X)\big)^2\right)}$. Then $\boxed{SD(X) = \sqrt{E(X^2) - \big(E(X)\big)^2}}$ **OR** $\boxed{SD(X) = \sqrt{E\left(\big(X - E(X)\big)^2\right)}}$. Usually, $E\big(f(X)\big) \neq f\big(E(X)\big)$.

	Properties: If X and Y are **independent**, then $\boxed{Var(X+Y)} = E(X+Y)^2 - \big(E(X)+E(Y)\big)^2 = E(X^2) + 2E(X)E(Y) - E(Y^2) - 2E(Y)E(X) = \boxed{Var(X)+Var(Y)}$. $\boxed{SD(aX+b) = \lvert a \rvert SD(X)}$ $\boxed{Var(aX+b) = a^2 Var(X)}$
Chebyshev Inequality	For any random variable X, and any $k > 0$, we have the Chebyshev inequality $\boxed{P\big(\lvert X - E(x)\rvert \geq k \cdot SD(x)\big) \leq \frac{1}{k^2} \Leftrightarrow P(\lvert X - E(x)\rvert \geq t) \leq \frac{Var(X)}{k^2}}$. Standardized Chebyshev is $P(X^* \geq k) \leq \frac{Var(X^*)}{t^*} \Rightarrow \boxed{P(\lvert X^* \rvert \geq t) \leq \frac{1}{t^2}}$.
Central Limit Theorem	If there is a random variable $X = \{X_1, X_2, \dots, X_n\}$, which every element is independent and identically distributed (iid). Then we have the following formula. $\boxed{P(a \leq X^* \leq b) = P(a \leq Z \leq b)}$, which is standard normal $N(0,1)$. As n goes to infinity, $P(a \leq X^* \leq b) = \Phi(b) - \Phi(a)$.