Definitions:

random experiment – an experiment whose individual outcomes are uncertain but there is a regular distribution in a large number of repetitions. Example: Coin tossing and dice rolling

outcome: the value of one replication of a random experiment. Written as little s. § Coin Tossing: s = H with one toss of a coin • s = HTT with three tosses

sample space (labeled S): is the set of all possible outcomes of a random experiment Toss a coin three times: S = {HHH,THH,HTH,...,TTT} Face showing when rolling a six-sided die: S = {1,2,3,4,5,6} Pick a real number between 1 and 20: S ={[1,20]}

event (labeled A, B, etc...): a set of outcomes of a random experiment. The event A that exactly two heads are obtained when a coin is tossed three times: A ={HHT,HTH,THH}

The **union** of two events A and B is the event that either A occurs or B occurs or both occur: C = (A or B) = (A **U** B)

The **intersection** of two events A and B is the event that both A and B occur. C = (A and B) = (A ∩ B) = AB

A implies B A ⊆ B (A is contained in B)

Naïve definition of probability: For a random phenomenon, if the sample space is finite and if all of the individual outcomes have the same probability, then the probability of an event A (written P(A)) is the ratio

$$P_{naive}(A) = \frac{\#\text{ of elements in } A}{\#\text{ of elements in } S} = \frac{|A|}{|S|}$$

A function that maps the outcomes in S (an experiment) to the real line is called a **random variable** (text, p.92), often written as r.v.

The **distribution** of X is the collection of all probabilities of the form P(X ∈ C) (means X is a member of C) for all sets C of real numbers such that {X ∈ C} is an event.

A random variable X is a **discrete random variable** if X can take only a finite number k of different values x1,..., xk or at most an infinite sequence of different x1, x2,... (aka, countably infinite). Examples: X = # heads in 3 flips of a coin.

The **indicator random variable** of an event A is the r.v. which equals 1 if A occurs and 0 otherwise. We will denote the indicator r.v. of A by IA or I(A).

Fundamental Bridge: Note that I(A) ~ Bern(p) with p = P(A).

$$P(A) = E(I_A)$$

For a discrete random variable X, the **probability mass function** (or simply just probability function) of X is defined as the function f such that for every real number x, f (x) = P(X = x)

A valid PDF: (1) It is always nonnegative. (2) The sum of its values, at all the places where it is nonzero, equals 1.

The **cumulative distribution function** of a r.v. X is the function F_X given by $F_X(x) = P(X \le x)$. It is often written as just capital F without the subscript, or F(x). (other letters, like G or H, can also be used).

CDFs are always non-decreasing and non-negative

limx→-∞[F(x)]= 0. limx→∞[F(x)]= 1.

CDFs are always continuous from the (right)

For a continuous r.v. X with CDF F, the **probability density function (PDF)** of X is the derivative, f, of the CDF, given by f(x) = F'(x): The support of X, and of its distribution, is the set of all x where f(x) > 0.

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

The kth **moment** (k is an integer) of a r.v. X is defined as the expectation $E(X^k)$.

The kth moment is said to exist if $E(|X^k|) < \infty$

The kth central moment (k is an integer) of a r.v. X is defined as the expectation E[(X-μ)k]

The **skewness** of a r.v. X with mean μ and variance σ² is the third standardized moment of X:

$$\text{Skew}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$, positive skew is a right-tailed distribution (skew for expo(lambda) is 2, and kurtosis is 6)

The kurtosis of a r.v. X with mean μ and variance σ² is the fourth standardized moment of X:

$$\text{Kurt}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$$

It is measuring how heavy tailed a distribution is. A prototypical distribution with large kurtosis has a PDF with a sharp peak in the center (within 1σ), low shoulders (±(1 to 2)σ), and heavy tails (beyond ±2σ). Unif has negative kurtosis.

Derivatives: $\frac{d}{dx}(x^n) = nx^{n-1}$, n is any number. $(fg)' = f'g + fg'$ – **(Product Rule)** $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$ – **(Quotient Rule)**

$\frac{d}{dx}\left(f(g(x))\right) = f'(g(x))g'(x)$ **(Chain Rule)** $\frac{d}{dx}(a^x) = a^x \ln(a)$ $\frac{d}{dx}(\mathbf{e}^x) = \mathbf{e}^x$

$\frac{d}{dx}\left(\mathbf{e}^{g(x)}\right) = g'(x)\mathbf{e}^{g(x)}$ $\frac{d}{dx}(\ln g(x)) = \frac{g'(x)}{g(x)}$ $\frac{d}{dx}(\ln(x)) = \frac{1}{x},\ x>0$ $\frac{d}{dx}(\ln|x|) = \frac{1}{x},\ x \ne 0$ $\frac{d}{dx}(\log_a(x)) = \frac{1}{x\ln a},\ x>0$

Integration:

u Substitution

Given $\int_a^b f(g(x))g'(x)dx$ then the substitution $u = g(x)$ will convert this into the integral, $\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)\,du$.

Integration by Parts

The standard formulas for integration by parts are,

$$\int udv = uv - \int vdu \qquad \int_a^b udv = uv\Big|_a^b - \int_a^b vdu$$

Choose u and dv and then compute du by differentiating u and compute v by using the fact that $v = \int dv$.

$\int x^n dx = \frac{1}{n+1}x^{n+1} + c,\ n \ne -1$

$\int x^{-1}dx = \ln|x| + c$ $\int x^{-n}dx = \frac{1}{-n+1}x^{-n+1} + c,\ n \ne 1$

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

$\int u\mathbf{e}^u du = (u-1)\mathbf{e}^u + c$

$\int \mathbf{e}^u du = \mathbf{e}^u + c$ $\int a^u du = \frac{a^u}{\ln a} + c$ $\int \ln u\, du = u\ln(u) - u + c$ $\int \frac{1}{u\ln u}du = \ln|\ln u| + c$

Probability Rules:

Inclusion-Exclusion:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &+ P(A \cap B \cap C) \end{aligned}$$

P(union of many events) = singles – doubles + triples – quadruples + ….

Counting:

Sampling table:

	Matters	Not Matter
With Replacement	n^k	$\binom{n+k-1}{k}$
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Binomial Coef:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)...(n-(k-1))}{k!}$$

Permutations:

The number of ways of arranging n objects, of which p of one type are alike, q of a second type are alike, r of a third type are alike, etc is:

$$\frac{n!}{p!\,q!\,r!\ \ldots}$$

Ex for STATISTICS: $\frac{10!}{3!\,2!\,3!}$ = 50,400

De Montmort's problem: Say the number on the random card. Solution: $1 - e^{-1} \approx 0.63$

Taylor Series for e^x:

$$\mathbf{e}^x = \sum_{n=0}^{\infty} \frac{1}{n!}x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Geometric Series:

$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$, as long as q < 1

Probability of birthday matches with k people in a room:

P(at least 1 birthday match) = 1 – P(no birthday matches)

$$= 1 - \frac{365 \cdot 364 \cdot 363 \cdots (365-(k-1))}{365^k}$$

Conditional Probability:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

P(A) ignoring B is often called the prior probability, and P(A|B) is often called the posterior probability incorporating knowledge/information of event B.

$$P(A|E) = 1 - P(A^C|E)$$

$$P(A \cup B|E) = P(A|E) + P(B|E) - P(A \cap B|E)$$

$$P(A|B \cap E) = \frac{P(A \text{ and } B|E)}{P(B|E)}$$

$P(A \cap B) = P(B|A)P(A) = P(A)P(B|A)$

$P(A \cap B \cap C) = P(A)\,P(B|A)P(C|A \cap B)$

Bayes Rule:

$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ or $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$

Equivalent definitions of Independence:

$P(A|B) = P(A)$

$P(A \cap B) = P(A)P(B)$

$P(A|B) = P(A|B^C)$

$P(B|A) = P(B)$

Two events are conditionally independent, given E, if:

$$P(A \cap B|E) = P(A|E)P(B|E)$$

Simpson's Paradox: Nick is better overall, but Hibbert is better at each category

	Heart	Band-Aid
Success	70	10
Failure	20	0

Dr. Hibbert

	Heart	Band-Aid
Success	2	81
Failure	8	9

Dr. Nick

Expectation:

The expected value (or expectation or mean), E(X), of a discrete random variable X is defined as

$$E(X) = \sum_{all\ x} xP(X = x)$$

The expected value of a continuous r.v. with pdf f(x) is

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

Expectation is linear, even if variables are dependent:

$$E(X+Y) = E(X) + E(Y)$$

$$E(aX+b) = aE(X) + b$$

LOTUS:

Discrete r.v.

$$E[g(X)] = \sum_{all\ x} g(x)P(X = x)$$

Continuous r.v.

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Variance:

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

Variance for a continuous r.v.

$$\text{Var}(X) = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Properties:

$\text{Var}(X + c) = \text{Var}(X)$

$\text{Var}(cX) = c^2\text{Var}(X)$

$\text{Var}(X) \ge 0$

If X and Y are independent:

$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Discrete r.v.s

Bernoulli Distribution:

Imagine you are going to flip a coin once, with known probability p of showing heads.
Let X = # heads.
If X~Bern(p), P(X = 1) = p and P(X = 0) = 1 – p

$E(X) = \sum_{all\ x} xP(X = x) = 0(1-p) + 1(p) = p$

MGF: M(t) = $q + pe^t$

Binomial Distribution: sum of iid Bernoulli trials.

Imagine you are going to flip a coin n times, with known probability p of showing heads each times. Assume the result of each flip is independent from one another.
Let X = # heads.
We write X ~ Bin(n, p) to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer (total number of trials) and 0 < p < 1.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$
, for k = 0,1,2,3,…,n

MGF: $M(t) = (q + pe^t)^n$

Hypergeometric Distribution:

Suppose we have an urn filled with w white and b black balls. Then drawing n balls out of the urn with replacement yields a Bin(n, p = w/(w+b)) distribution for the number of white balls obtained in n trials, since the draws are independent Bernoulli trials, each with probability p = w/(w+b) of success.
If we instead **sample without replacement**, then X = the number of white balls follows a Hypergeometric distribution.
X ~ HGeom(w, b, n), then the PMF of X is

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$
, for integers k satisfying 0 ≤ k ≤ w and 0 ≤ n – k ≤ b.

In a five-card hand drawn at random from a well-shuffled standard deck, the number of aces in the hand has a HGeom(4, 48, 5) distribution, which can be seen by thinking of the aces as white balls and the non-aces as black balls.
If X ~ Bin(n,p) and Y ~ Bin(m,p) and X, Y are independent, then the conditional distribution of X given X+Y = r is Hgeom(n,m,r)

Discrete Uniform Distribution:

Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X. Then X is said to have the Discrete Uniform distribution with parameter C; we will denote this as X ~ DUnif(C).
The PMF of X ~ DUnif(C) is:

$$P(X = x) = \frac{1}{|C|}$$

for x ∈ C (and 0 otherwise).

Geometric Distribution:

Consider a sequence of independent Bernoulli trials, each with the same success probability p, with trial performed until a success occurs. Let X be the number of failures before the first successful trial. Then X has the Geometric distribution with parameter p; we denote this by X ~ Geom(p)
X ~ Geom(p), then the PMF of X is:
$P(X = k) = (1-p)^k p = q^k p$
for k = 0, 1, ...
E(X) = (1-p)/p
Var(X) = (1-p)/p^2
MGF: $M(t) = \left(\frac{p}{1-qe^t}\right)^r$
Proof the Geometric dist is Memoryless:

$$P(X > s+t | X > t) = \frac{P((X > s+t) \cap (X > t))}{P(X > t)} = \frac{P(X > s+t)}{P(X > t)} = \frac{(1-p)^{s+t}}{(1-p)^t} = (1-p)^s$$

Negative Binomial Distribution:

In a sequence of independent Bernoulli trials with success probability p, if X is the number of failures before the rth success, then X is said to have the Negative Binomial distribution with parameters r and p, denoted X ~ NBin(r, p)
* Also: sum of iid geometric(p) rv's
If X ~ NBin(r, p), then the PMF of X is:

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

for n = 0, 1, 2, ...

	With replacement	Without replacement
Fixed number of trials	Binomial	Hypergeometric
Fixed number of successes	Negative Binomial	Negative Hypergeometric

Poisson Distribution:

Imagine you are trying to determine the number of occurrences ("successes") of a certain rare type of cancer (melanoma) in a large population (like the state of Massachusetts) over a fixed period of time (say a year). The Poisson distribution is instead often used in situations like this, where we are counting the number of successes in a particular region or interval of time, and there are a large number of trials, each with a small probability of succes
$X \sim \mathrm{Pois}(\lambda)$, if the PMF of X is:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

for $k = 0, 1, \ldots$ This is a valid PMF b/c taylor series: $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$

The parameter λ is interpreted as the rate of occurrence of these rare events; in the examples above, could be 20 (emails per hour) or 10 (chips per cookie)
E(X) = Var(X) = λ

Negative Hypergeometric:

An urn with b black balls and w white balls, where balls are drawn 1 by 1 without replacement.
X = number of black balls drawn before any white balls.
E(X) = b/(w +1)

Continuous r.v.s

Uniform Distribution:

Consider a completely random number (with real value) between the values a and b, each with equal likelihood.
Let the r.v. X be the value of this completely random number on the interval (a,b).
Then X has the Uniform distribution with parameters a and b; we denote this by X ~ Unif(a,b)

PDF: $f(x) = \frac{1}{b-a}, \ a < x < b$

$E(X) = (a+b)/2$

$\mathrm{Var}(X) = (b-a)^2/12$

CDF: (x-a)/(b-a) for x in [a,b]

Universality of the Uniform:

1. Let U ~ Unif(0,1) and X = F^{-1}(U). Then X is an r.v. with CDF F.
2. Let X be an r.v. with CDF F. Then F(X) ~ Unif(0, 1).

Normal Distribution:

Because of the central limit theorem which says that under very weak assumptions, the sum (or average) of a large number of i.i.d. random variables has an approximately Normal distribution, regardless of the distribution of the individual r.v.s.

Pdf: $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ -\infty < x < \infty$$

To Convert to standard: Let Z ~ N(0,1), and let X = μ + σZ. Then X ~ N(μ, σ^2).

MDF: $M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{(\sigma t)^2/2} = e^{\mu t + (1/2)\sigma^2 t^2}$

Standard Normal Distribution:

Pdf: $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \ -\infty < x < \infty$

CDF: $$\Phi(z) = \int_{-\infty}^{z} \varphi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Emperical Rule:
$P(|X - \mu| < \sigma) \approx 0.68$
$P(|X - \mu| < 2\sigma) \approx 0.95$
$P(|X - \mu| < 3\sigma) \approx 0.997$

Exponential Distribution:

Consider you are waiting (in continuous time) until a success for some process/experiment occurs, where λ is the average # successes per unit of time.
Let X be the amount of time you have to wait until this success arrives.
Then X has the Exponential distribution with parameter λ; we denote this by X ~ Expo(λ)
The Exponential distribution is the continuous counterpart to the Geometric distribution
If X ~ Expo(λ), then the PDF of X is (λ > 0):

$$f(x) = \lambda e^{-\lambda x}, \ x > 0$$

The corresponding CDF is:

$$F(x) = 1 - e^{-\lambda x}, \ x > 0$$

Exponential Scaling:
If X ~Expo(1), then

$$Y = \frac{X}{\lambda} \sim \mathrm{Expo}(\lambda)$$
, also if Y~Expo(λ), then Yλ ~ Expo(1)

MGF if Y~Expo(λ): $M_Y(t) = M_X(t/\lambda) = \frac{1}{1-t/\lambda} = \frac{\lambda}{\lambda - t}$

If X~ Expo(1) MGF $E(X^n) = n!$

Memoryless property:
A distribution is said to have memorylessness property if a random variable X from that distribution satisfies:

$$P(X \ge s+t \mid X \ge s) = P(X \ge t)$$

Proof that Expo(λ) is memoryless:
If X~Expo(λ), P(X≥s+t|X≥s) = P(X≥s+t)/ P(X≥s) = e^(- λ(s +t)) / e^(- λs) = e^(- λt) = P(X>=t)

Poisson Process:
A process of arrivals in continuous time is called a Poisson process with rate λ if the following two conditions hold:
1. The number of arrivals that occur in an interval of length t is a Pois(λt) random variable.
2. The numbers of arrivals that occur in disjoint intervals are independent of each other. For example, the numbers of arrivals in the intervals (0; 10); [10; 12); and [15;19) are independent.
The time until the 1st arrival in a Poisson process of rate λ has an Exponential distribution with parameter λ

* min of exponentials is ~Expo(λ1, λ2)

Mean, Median, Mode:

We say that c is a **median** of a r.v. X if P(X ≤ c) ≥ 1/2 and if P(X ≥ c) ≥ 1/2
For a discrete r.v. X, we say that c is a **mode** of X if it maximizes the PMF: P(X = c) ≥ P(X = x) for all x. For a continuous r.v. X with PDF f, we say that c is a **mode** if it maximizes the PDF: f (c) ≥ f (x) for all x

The value of c that minimizes the mean squared error,
E[(X – c)²], is c = μ.
A value of c that minimizes the mean absolute error,
E(|X – c|), is c = m

###################### may remove
Let X1,...,Xn be i.i.d. random variables with mean μ and variance σ². Then Xn is unbiased for estimating μ. That is:

$$E(\bar{X}_n) = \mu$$
and variance of the sample mean is
$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

Sample variance is $$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$
############################

Moments and Moment Generating Functions:
The moment generating function (MGF) of a r.v. X is M(t) = E(e^{tX}), as a function of t, if this is finite on some open interval (-a, a) containing 0. Otherwise, the MGF of X does not exist.
M (t) = E(e^{tX})
Let X ~ Geom(p). Calculate the MGF for X.
$M(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} q^k p = p\sum_{k=0}^{\infty} (qe^t)^k = \frac{p}{1-qe^t}$, and t can be numbers near zero.

Let U ~ Unif(a,b). Calculate the MGF for U

$$M(t) = E(e^{tX}) = \int_a^b e^{tu}\left[\frac{1}{b-a}\right]du = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Given the MGF of X we can get the nth moment of X by evaluating the nth derivative of the MGF at 0:

$$E(X^n) = M^{(n)}(0)$$

For example, we can find E(X) for the Geometric(p) dist.:

$$M(t) = \frac{p}{1-qe^t} \Rightarrow M'(t) = \frac{pqe^t}{(1-qe^t)^2}$$

$$\Rightarrow E(X) = M'(0) = \frac{pqe^0}{(1-qe^0)^2} = \frac{pq}{(1-q)^2} = \frac{q}{p}$$

If two r.v.s have the same MGF, they must have the same distribution.

If X and Y are independent, then the MGF of X + Y is the product of the individual MGFs:
$$M_{X+Y}(t) = M_X(t)M_Y(t)$$
This is true because if X and Y are independent, then:
$$E(e^{t(X+Y)}) = E(e^{t(X)}e^{t(Y)}) = E(e^{t(X)})E(e^{t(Y)})$$

Location-scale transform MGF:
$$E(e^{t(a+bX)}) = e^{ta}E(e^{btX}) = e^{at}M(bt)$$

Problems:
Number of ways to split 360 people into 120 teams of 3.

$$\frac{360!}{((3!)^{120} * 120!)}$$

360! Is the ways to order 360 ppl
3!^120 is the number per team, 120! Is the ways to order 120 teams.