|  |  |  |  |
| --- | --- | --- | --- |
| 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 **** 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  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).    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 FX given by FX(x) = P(X ≤ 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.    The kth **moment** (k is an integer) of a r.v. X is defined as the expectation E(Xk).  The kth moment is said to exist if E(|Xk|) < ∞  **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:  , 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:  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:    Integration:        Logarithm: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-22 at 9.31.01 PM.png  **Probability Rules:**  Inclusion-Exclusion:    P(union of many events) = singles – doubles + triples – quadruples + ….  **Counting:**  Sampling table:    Binomial Coef:    **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:  n!        p! q! r! …  Ex for STATISTICS: 10! = 50,400  3! 2! 3!  De Montmort’s problem: Say the number on the random card. Solution: | |
| **Taylor Series for e^x:**    **Geometric Series:**  , as long as q < 1  Probability of birthday matches with k people in a room:    **Conditional Probability:**    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. | **Bayes Rule:**  or  Equivalent definitions of Independence:      Two events are conditionally independent, given E, if:    **Simpson’s Paradox:** Nick is better overall, but Hibbert is better at each category | | **Expectation:**  The expected value (or expectation or mean), E(X), of a discrete random variable X is defined as    The expected value of a continuous r.v. with pdf f(x) is    Expectation is linear, even if variables are dependent:    **LOTUS:**  Discrete r.v.    Continuous r.v.    Variance:  **or**  Variance for a continuous r.v.    Properties:    If X and Y are independent: |

|  |  |  |
| --- | --- | --- |
| **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    MGF: M(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.  , for k = 0,1,2,3,…,n  MGF:    **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  , 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)  \*Sum of hypergeometric distn’s is not hypergeometric!  **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:    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:    for k = 0, 1, ...  E(X) = (1-p)/p  Var(X) = (1-p)/p^2  MGF:  Proof the Geometric dist is Memoryless:  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-22 at 6.56.26 PM.png  **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:    for n = 0, 1, 2, ...    **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  This is a valid PMF b/c taylor series:  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:**    **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:  To Convert to standard: Let Z ~ N(0,1), and let X = μ + σZ. Then X ~ N(μ, σ2).  MDF:  **Standard Normal Distribution:**  **Pdf:**  CDF:  Emperical Rule:    Normal Transformations:  If X~N(mu, sigma^2),  **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):    The corresponding CDF is:    Exponential Scaling:  If X ~Expo(1), then  , also if Y~Expo(λ), then Yλ ~ Expo(1)  MGF if Y~Expo(λ):  If X~ Expo(1) MGF  Memoryless property:  A distribution is said to have memorylessness property if a random variable X from that distribution satisfies:    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:  and variance of the sample mean is  Sample variance is  ###########################  Moments and Moment Generating Functions:  The moment generating function (MGF) of a r.v. X is M(t) = E(etX), 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(etX )  Let X ~ Geom(p). Calculate the MGF for X.  , and t can be numbers near zero.  Let U ~ Unif(a,b). Calculate the MGF for U    Given the MGF of X we can get the nth moment of X by evaluating the nth derivative of the MGF at 0:    For example, we can find E(X) for the Geometric(p) dist.:    **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:    This is true because if X and Y are independent, then:    Location-scale transform MGF:    Problems:  Number of ways to split 360 people into 120 teams of 3.  360! Is the ways to order 360 ppl  3!^120 is the number per team, 120! Is the ways to order 120 teams.  Sums of independent r.v.s |

Expectation

E(Y1 + Y2|X) = E(Y1|X) + E(Y2|X) (Linearity);

E(Y |X) = E(Y ) if X and Y are independent;

E(h(X)Y |X) = h(X)E(Y |X) (Taking out what’s known);

E(Y ) = E(E(Y |X)) (Iterated Expectation/Adam’s Law);

Var(Y ) = E(Var(Y |X)) + Var(E(Y |X)) (Eve’s Law).

Chicken-Egg story:

A chicken lays a Poisson(λ) number N of eggs. Each egg, independently,

hatches a chick with probability p. Let X be the number which hatch, so

X|N ⇠ Bin(N,p).

As shown in class, in this story X is independent of Y , with X ⇠ Pois(λp) and

Y ⇠ Pois(λq), for q = 1 – p

Convert gamma to Chi Sq:

If X ~ gamma(k,b) then Y=2bX has the chi-square distribution with 2k degrees of freedom.

Gamma(1/2, ½) ~ chiSq(1)

Series:

