Distribution overall pattern of how often values occur **Distribution Properties** shape (*symmetric* or *skewed*), location, variability, deviations **Location** center—typically measured by *mean* (avg/expected val) or median **Variability** typically measured by *variance* or *SD*, or range **Deviations** from overall pattern, e.g. possible outliers, or unusual points that are not consistent with rest of data **Stem-and-Leaf Plots** leading digits for *stem*, trailing digits for *leaves*; list *stem* in a vertical column; record the *leaf* for each obs right to the column of the *stem*; indicate unit **Histogram** more suitable for big data sets **Sample Mean** $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ for $x_1 ... x_n$ **Sample Mean as Location** may not be representative; more robust would be **Sample Median as Location** middle value of a data set; robust (not sensitive) to outliers **Sample Median** for odd n, $\tilde{x} = x_{(n+1)/2}$ for even n, $\tilde{x} = \frac{x_{(n/2)}+x_{(n/2+1)}}{2}$ **Sample $\bar{x}$, $\tilde{x}$ vs Population** sample mean is "estimate" of population mean (μ); sample median is "estimate" of population median. Estimation improves as sample size grows **($\bar{x}$, $\tilde{x}$) for Distro Shapes** if symmetric, mean=median if right-skewed, mean>median if left-skewed, mean<median **Variability as Sample Variance** $s^2 = \frac{\sum_{i=1}^{n}(x_i^2) - n^{-1}(\sum_{i=1}^{n}(x_i))^2}{n-1}$ **Sample Standard Deviation** $s = \sqrt{s^2}$ **Sample S.D. Properties** S has same unit as data—can be interpreted as *representative deviation* of data from center **Sample Variance & S.D. vs Population** sample variance S^2 is an "estimate" of population variance (σ^2); sample S.D. s is an "estimate" of population SD (σ). Estimation improves as sample size grows **Experiment** any action or process whose outcome is subject to uncertainty (e.g. flipping a coin, rolling a die) **Sample Space** set of all possible outcomes of an experiment (ς) e.g. flipping one coin, ς =(H,T) *order of outcomes matters!* **Event** any collection of outcomes from the sample space (usually denoted by a capital letter) **Complement** of event A: set of all outcomes ς that are not in A (A') **Intersection** $A = \{1,2,3\}$, $B = \{1,3,5\}$, $A \cap B = \{1,3\}$ **Union** A={1,2,3}, B={1,3,5}, $A \cup B = \{1,2,3,5\}$ **Mutually Exclusive/Disjoint Events** if $A \cap B = \varnothing$ **Important Results** $A \cap \varnothing = \varnothing$, $A \cup \varnothing = A$, $A \cap A' = \varnothing$, $A \cup A' = \varsigma$, $\varsigma' = \varnothing$, $\varnothing' = \varsigma$, $(A')' = A$, $(A \cap B)' = A' \cup B'$, $(A \cup B)' = A' \cap B'$ **Probability** precise measure of the chance that a particular event will occur **P(A)** probability of event A occurring e.g. $P(H) = 0.5$ **P(A) for Equally Likely Events** if outcomes in ς are equally likely to occur, $P(A) =$ (number of outcomes in event A)/(number of outcomes in ς) **Probability as Long-Run Average** probability represents the proportion or percent of the time (over an extended period) we would expect an event to happen (more or less likely, from historical data) **Probability and Sample Size** probability estimates based on *small sample sizes* are not as reliable as those based on *large sample sizes* **Core Properties of Probability** $P(A) \geq 0$, $P(S) = 1$, $P(A_1 \cup A_2 \cup ...) = \sum_{i=1}^{\infty} P(A_i)$ for infinite series of mutually exclusive events A_1... or for finite series of events **Additional Properties** $P(A) = 1 - P(A')$, $P(A) \leq 1$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$ **Conditional Probability** probability of A, given B: calculated with $P(A|B) = \frac{P(A \cap B)}{P(B)}$ if $P(B) > 0$ **Key Idea** is that what will happen when *something is known* and is thus *no longer uncertain*; how often A will occur given knowledge that B *has occurred* **Drug Test Example** $P(A|B) =$ given that the person used the drug, probability that (s)he tested positive: *sensitivity*, $P(B'|A') =$ given that person did not use, probability that (s)he tested negative: *specificity* **Probability Product Rule** $P(A \cap B) = P(A|B) \cdot P(B)$ **Tree Diagram** see end **Video Game Example** $P(W_2) = P(W_1 \cap W_2) + P(L_1 \cap W_2)$ **Law of Total Probability** let $A_1, ..., A_k$ be mutually exclusive, exhaustive events. For any event B, $P(B) = P(B|A_1)P(A_1) + ... + P(B|A_k)P(A_k) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$ **Using LoTP for Events** LoTP allows us to calc probabilities by conditioning on other events. Sometimes it's easier to find $P(B|A)$ than $P(B)$. But if it's easier to find $P(B|A)$ than $P(A|B)$, use **Bayes' Theorem** for mutually exclusive and exhaustive events $A_1, ..., A_k$ with prior probabilities $P(A_i) > 0 (i = 1, ..., k)$, for any event B for which $P(B) > 0$, the probability of event A_j given that B has occurred is $P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$, $j = 1, ..., k$ **Bayes' Thm Properties** numerator uses multiplication (product) rule; denominator uses LoTP; Bayes' Thm is application of both together **Video Game Example** Given that you won your second play, probability that you won your first play is $P(W_1|W_2) = \frac{P(W_2|W_1)P(W_1)}{P(W_2|W_1)P(W_1)+P(W_2|L_1)P(L_1)}$ **Drug Test Example** probability that person took drug, given positive test, is $P(D|T+) = \frac{P(T+|D)P(D)}{P(T+|D)P(D)+P(T+|ND)P(ND)}$ **Independence I** Two events A and B are *independence I* iff $P(A|B) = P(A)$; otherwise the events are *dependent* **Temperature Example** A = Raleigh temp, B = Paris temp. A,B independent bc knowing Paris temp does not help predict Raleigh temp. If C=tomorrow's Durham high, A and C are dependent because *predictions about Raleigh's temp are surely affected by knowing Durham's temp* **Independence II** A and B are independent iff $P(A \cap B) = P(A) \cdot P(B)$ **V.G. Example** $P(W_2|W_1) = P(W_2)$, so W_1 and W_2 are *independent*; also, $P(W_2 \cap W_1) = P(W_1)P(W_2)$, so W_1 and W_2 **Independence for Many Events** events $A_1, ..., A_n$ are mutually independent if, for every $k (k = 2, 3, ..., n)$, and every subset of indices $i_1, i_2, ..., i_k$, $P(A_{i_1} \cap A_{i_2} \cap ... A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$ **M.C. Test** 20 q's, 4 possible answers; student has 80% chance of being correct. q's are independent. Let A_j = student's answer is correct on question j. Events are independent. Then $P(\text{perfect score}) = P(A_1 \cap A_2 \cap \cdots \cap A_{20}) = P(A_1) \cdot P(A_2) \cdots P(A_{20}) = \prod_{j=1}^{20} 0.8 = 0.012$ **Circuit System Example** see end **Monty Hall Problem** before door opened, w/ H3 = host opened 3: no car, $P(C_1) = P(C_2) = P(C_3) = 1/3$. Knowing H_3, $P(C_1|H_3) = \frac{P(C_1 \cap H_3)}{P(H_3)} = \frac{P(H_3|C_1)P(C_1)}{P(H_3)}$. C_1, C_2, C_3 are mutually exclusive and exhaustive: $C_1 \cup C_2 \cup C_3 = \varsigma$. Therefore, $P(C_1|H_3) = \frac{P(H_3|C_1)P(C_1)}{P(H_3|C_1)P(C_1)+P(H_3|C_2)P(C_2)+P(H_3|C_3)P(C_3)}$. We know $P(H_3|C_3) = 0$, $P(H_3|C_2) = 1$ (you picked door 1), $P(H_3|C_1) = 1/2$ (since host randomly picks from door 2 or 3). Thus $P(C_1|H_3) = P(\text{win if stay}) = 1/3$, while $P(C_2|H_3) = P(\text{win if switch}) = 2/3$. **Data Types** *quantitative* (numbers) or *qualitative* (words) **Random Variable (rv)** a function that associates each element of sample space with a number **rv Notation** usually denoted by capital letters, e.g. X,Y,Z, specific values denoted by lower case, e.g. x,y,z; $P(Y = y) =$ prob that rv Y equals the value y **Opinion Example** for ς ={Strongly agree, agree, disagree, strongly disagree}, define X s.t. $X(SA) = 1, S(A) = 1, X(SD) = 0, X(D) = 0$ **rv and ς** rv's are a function with an input that is an element of the sample space and an output that is a $\Re$; any characteristic whose value can change over the sample space **Discrete** number of possible values is finite or countably infinite **Discrete rv Examples** $X = 1$ if male, 0 if female (finite values); $X = \#$ of additional coin flips before 1st tails is obtained (countable: 0,1,2,...)) **Continuous** X *continuous* X is continuous if (i) X can be any value in an interval, such as $[0,1]$ or even $(-\infty, +\infty)$, or a union of disjoint intervals, AND (ii) $P(X = c) = 0$ for any possible value of c—i.e. no individual val of X has a positive probability **Probability Distro (pmf)** of a discrete rv: Describes how likely the possible values of x (i.e. possible inputs) are to occur. **pmf Notation** $p(x) = P(X = x)$ **Conditions for pmf** $p(x) \geq 0$ for any x; sum of $P(X = x)$ over all possible x must be 1. Anything that satisfies these is a valid pmf. **pmf Table** see end. **pmf Parameters** some pmf's are indexed by parameters, which are quantities that can take any one of several possible values. Each possible value of a parameter defines a different pmf. **Parameter Example** $p(x) = \alpha$ if $X = 1$, $1-\alpha$ if $X = 0$, for $\alpha \epsilon (0,1)$. If $\alpha = 0.5$, $P(X = 1) = 0.5 = P(X = 0)$ **Family of Distros** set of all pmf's obtained by varying a parameter **Family Properties** families of distros all share a common function; not every family/function has a formal name **Bernoulli Family** $p(x) = \alpha$ if $X = 1$, $1 - \alpha$ if $X = 0$, for $\alpha \in (0,1)$ $X\sim$ **Bernoulli(α)** X is distributed as Bernoulli with parameter α **Cumulative Distribution Function (cdf)** of a drv X: provides probability that the observed value of X will be at most x: $F(x) = P(X \le x) = \sum_{y:y\le x} p(y)$ (sum over all values less than equal to x **Expected Value** given drv X, w/ set D of possible values and pmf $p(x)$. The *expected value* or *mean* of X is: $E(X) = \mu_X = \sum_{x\in D} x \cdot p(x)$, where $E(X)$ stands for Expectation of X, weighted average of all its possible values with weights being probabilities. **Expected Value Properties** expected val is most commonly used measure of central tendency; it is a single number that gives some info about entire distribution. Also, a "typical" value for rv (tho not necessarily most common, or even a val rv can take); balancing point: mean value such that teeter-totter distro would balance; long-run signal for noisy process. Can be used to compare distros **Expected Value of a Function** for drv X, with set D of possible values, pmf $p(x)$, and function $h(X)$. Then expected value of $h(X)$ is $E[h(X)] = \sum_D h(x) \cdot p(x)$. $h(X)$ is another rv, with a dist. But we don't need the dist to find $E[h(X)]$. **Bernoulli e^x Example** given $p(x) = \alpha$ if $X = 1$ and $1 - \alpha$ if $X = 0$, what is expected val of $h(X) = e^X$? It's $E(e^X) = \sum_{x=0}^{1} e^x \cdot p(x) = e^0 \cdot p(0) + e^1 \cdot p(1) = 1 \cdot (1-\alpha) + e \cdot \alpha = e\alpha - \alpha + 1$ $E(a + bX)$ $= a + bE(X) = a + b\mu$ where $\mu = E(X)$ **Temp Example** if X = temp in F and Y = temp in C = $(5/9)(X - 32)$, and if $E(X) = 71$F, then $E(Y) = E[\frac{5}{9}(X-32)] = \frac{5}{9}[E(X)-32] = \frac{5}{9}[71-32]$ **Variance** of drv X with set of possible values D, pmf $p(x)$, and mean μ is $V(X)$ $= \sigma_X^2 = \sum_D (x-\mu)^2 = E[(X-\mu)^2]$ **Standard Deviation** is square root of variance, or $\sigma_X = \sqrt{\sigma_X^2}$ **Notes on σ_X^2 and σ_X** both variance and SD are measured of the *spread* of a distribution. Variance has squared units; SD has same units as the rv! Due to this, SD is reported in practice. **σ_X^2, σ_X, and μ** σ_X^2 represents "expected square deviation from mean"; σ_X is approx *avg distance of observations from mean.* SD, with μ, gives better understanding of data values: *mean $\pm$ SD gives range of "typical" values for variable, better than just mean could* **Variance Equations** Variance can be represented in terms of expectations: $V(X) = E(X^2) - \mu^2$, while, for a function $a+bX$, variance is $V(a+bX) = b^2V(X)$ **Variance Eqn Properties** adding a constant to X doesn't affect its variance: $V(a+X) = V(X)$; multiplying X by a constant multiplies the SD by the magnitude of that constant: $V(bX) = b^2V(X)$, implies $\sigma_{bX} = |b|\sigma_X$ **Table Example** find mean and variance for the rv X and rv $Y = 3 + 7X$. When possible, use shortcut formulae. If $P(X = x) = 0.2$ for $X = 7$, 0.6 for $X = 12$, and 0.2 for $X = 14$, then $E(X) = 7(0.2) + 12(0.6) + 14(0.2) = 11.4 = \mu_X$; $V(X) = E(X^2) - \mu_X^2 = 135.4 - 11.4^2 = 5.44 = \sigma_X^2$, where $E(X^2) = 7^2(0.2) + 12^2(0.6) + 14^2(0.2)$. $E(Y) = 3 + 7\mu_X = 3 + 7(11.4) = 82.8 = \mu_Y$, and $V(Y) = 7^2\sigma_X^2 = 7^2(5.44) = 266.56 = \sigma_Y^2$ **Combinations** unordered groups of size r that can be formed from the n individuals in a group is "n choose r", or $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ **Example** $\binom{20}{18} = \frac{20!}{18!2!} = 190$ **Test Example** Test with 20 unrelated q's, each worth one point. $X_j = 1$ if student is correct on question j, 0 otherwise. Assume $X_j\sim$Bernoulli are independent: i.e., pmf is $p(x) = p$ if $X_j = 1$, $(1-p)$ if $X_j = 0$ for *each* of the rv's. $P(perfectscore) = P(X_1 = 1 \cap X_2 = 1 \cap ... \cap X_{20} = 1) = P(X_1 = 1) \cdot P(X_2 = 1) \cdots P(X_{20} = 1) = p^{20}$. If $p = 0.8$, $P(perfectscore) = (0.8)^{20} = 0.012$. $P(18correct)$ calculated in prior example. $P(18correct) = (190)(p)^{18}(1-p)^2$ **Binomial pmf** $\binom{n}{x} p^x(1-p)^{n-x}$ **Binomial pmf Notation** $X\sim$Bin(n,p) or b$(x;n,p)$ where X can take values $0, 1, 2, ..., n$ ranging from all failures to all successes. **Binomial pmf Props** if $n = 1$, Binomial$(1,p)$ =Bernoulli(p). Only 1 trial, X can only be 0 or 1, so we have $p(x) = \binom{1}{1}p^1(1-p)^{1-1} = p$ if $X = 1$, $\binom{1}{0}p^0(1-p)^{1-0} = (1-p)$ if $X = 0$ **Conditions for Using BinDist** (1) There are two possible outcomes of each trial: "success" (what we are counting) and "failure" (everything else (2) # trials n is known and fixed (3) outcomes are independent from one trial to others (4) probability of "success" p is the same for all trials. If conditions met, X ={number of total successes} is a binomial rv with parameters n =number of trials/sample size, p =the probability of success **Mean of $X\sim$Bin(n,p)** is **E(X)=np**, while the **Variance of $X\sim$Bin(n,p)** is $V(X) = np(1-p)$ **Maximizing $V(X)$** $V(X)$ is largest (given n) when $p = 0.5$ *because the outcome is hardest to predict!* **$V(X)$ Graphically** see end **Applying Binomial Formula** if $X\sim$Bin$(n = 3, p = 0.1)$, what is $P(X = x)$? $P(X = 0) = \binom{3}{0}(0.1)^0(1-0.1)^{3-0} = 0.729$, while $P(X = 1) = \binom{3}{1}(0.1)^1(1-0.1)^{3-1} = 0.243$, $P(X = 2) = \binom{3}{2}(0.1)^2(1-0.1)^{3-2} = 0.027$, and $P(X = 3) = \binom{3}{3}(0.1)^3(1-0.1)^{3-3} = 0.001$ (3.73) If A and B are independent events, show that A' and B are also independent. By LoTP, $P(B) = P(A \cap B) + P(A' \cap B)$. Then $P(A' \cap B) = P(B) - P(A)P(B)$ (by independence of A and B) $= [1 - P(A)]P(B) = P(A')P(B)$ **Ancestors protect me... :)**