Special Distributions (Discrete RVs)

E and Var	NAME	R_X	PMF
p & p(1-p)	Bernoulli(p)	{0,1}	p for x=1, 1-p for x=0
1/p and (1-p)/p^2	Geometric(p)	Z^+	$p(1-p)^{(k-1)}$ for k ∈ Z^+
np and np(1-p)	Binomial(n,p)	{0,1,..,n}	$^nC_k \cdot p^k \cdot (1-p)^{(n-k)}$ for k = 0 to n
m/p and (m.(1--p))/p^2	Pascal(m,p)	{m,m+1, m+2...}	$^{(k-1)}C_{(m-1)} \cdot p^m \cdot (1-p)^{(k-m)}$ for k = m, m+1, m+2, m+3, ...
np and ((b+r--n)/(b+r-1)).n-p(1-p)	Hypergeometric(b,r,k)	{max(0,k-r), max(0,k-r)+1, ..., min(k,b)}	$(^bC_x \cdot {}^rC_{(k-x)})/(^{(b+r)}C_k)$ ∀ x ∈ R_X
Both equal to lambda	Poisson(lambda)	Z^+	$(e^{-lambda} \cdot lambda^k)/k!$ for k ∈ R_X

Continuous RVs, PDFs and Mixed RVs

RV X with CDF $F_X(x)$ is continuous if $F_X(x)$ is a continuous function ∀ x ∈ R

PMF doesn't work for CRVs, since ∀ x ∈ R, $P_X(x)$ = 0. Instead, PDFs are used.

PDF = $f_X(x)$ = $dF_X(x)/dx$ (if $F_X(x)$ is differentiable at x) >=0 ∀ x ∈ R.

P(a<X<=b) = integral from a to b ($f_X(u)$.du) and **integral from -inf to +inf ($f_X(u)$. du) = 1**

Continuous RVs, PDFs and Mixed RVs (cont)

EX = integral from -inf to +inf (x . $f_X(x)$. dx) and **E[g(X)] = integral from -inf to +inf (g(x) . $f_X(x)$. dx)**

Var(X) = integral from -inf to +inf (x^2 . $f_X(x)$. dx - mu^2_X)

If g: R-> R is strictly monotonic and differentiable, then **PDF of Y=g(X) is $f_Y(y)$ = $f_X(x_1)$. $|dx_1/dy|$ where $g(x_1)$=y and 0 if g(x) = y has no solution**

Joint Distributions: RVs >= 2

Joint PMF of X and Y = $P_{XY}(x,y)$ = P(X=x, Y=y) = P((X=x) and (Y=y)) and **Joint range = R_{XY} = {(x,y)| $P_{XY}(x,y)$ > 0}** and **summing up P_{XY} over all (x,y) pairs will result in 1**

Marginal PMF of X = $P_X(x)$ = sum over all y_j ∈ R_Y ($P_{XY}(x, y_j)$) for any x ∈ R_X. Similarly, **Marginal PMF of Y = $P_Y(y)$ = sum over all x_i ∈ R_X ($P_{XY}(x_i, y)$) for any y ∈ R_Y**

To show independence between X and Y, **prove P(X = x, Y=y) = P(X=x) . P(Y=y) for all x-y pairs**. Similarly, **for conditional independence, show that P(Y=y|X=x) = P(Y=y) for all x-y pairs**

Joint CDF = $F_{XY}(x,y)$ = P(X<=x, Y<=y) and **Marginal CDF for X = $F_X(x)$ = limit y to inf($F_{XY}(x,y)$) for any x** and **Marginal CDF for Y = $F_Y(y)$ = limit x to inf ($F_{XY}(x,y)$) for any y**

Conditional expectation: E[X|Y=y_j] = sum over all x_i ∈ R_X (x_i . $P_{X|Y}(x_i|y_j)$)

NOTE: F_{XY}(inf,inf) = 1, F_{XY}(-inf,y) = 0 for any y and F_{XY}(x,-inf) = 0 for any x

P(x_1 < X <= x_2, y_1 < Y <= y_2) = $F_{XY}(x_1,y_1)$ + $F_{XY}(x_2,y_2)$ - $F_{XY}(x_2,y_1)$ - $F_{XY}(x_1,y_2)$

Conditional PMF given event A = $P_{X|A}(x_i)$ = P(X=x_i|A) = P(X=x_i and A)/P(A) for any x_i ∈ R_X and **Conditional CDF = $F_{X|A}(x)$ = P(X <= x | A)**

Joint Distributions: RVs >= 2 (cont)

Given RVs X and Y, $P_{X|Y}(x_i, y_j)$ = $P_{XY}(x_i,y_j)/P_Y(y_j)$. Similarly for Y|X

E[X + Y] = E[X] + E[Y] - independence not required

E[X . Y] = E[X] . E[Y] - independence IS required

Problem Solving Techniques

*** CARD PROBLEMS:**

Number of ways to pick k suits = ^4C_k with k=1,2,3,4

*** n BALLS, r BINS:**

- Distinguishable balls: each ball can go into any 1 of r bins. The # of distinct perms would be ^rP_n = r^n

- Indistinguishable balls: there will be 2 cases:

* No empty bins. Occupancy vector is x_1+...+x_r=n where every x is >= 1. There can be n-1 possible locations for bin dividers from which we can choose r-1 to keep >= 1 ball in each bin. # of possible arrangements = $^{(n-1)}C_{(r-1)}$.

* Bin may have 0 balls. Then the occupancy vector would be y_1+...+y_r=n+r and the # of arrangements will be $^{(n+r-1)}C_{(r-1)}$

*** COMMITTEE SELECTION: Solve using product rule/hypergeometric approach.**

*** HAT MATCHING PROBLEM:**

> Probability of k men drawing their own hats (over all k-tuples) = **(^nC_k(n-k)!)/n! = 1/k!**

# of derangements = **n![1-1/1!+1/2!--1/3!+...+$(-1)^n$/n!]**

> P(k matches) = **[1/2! - 1/3! + 1/4! -...+$(-1)^{(n-k)}$/(n-k)!]/k!**

Problem Solving Techniques (cont)

*** DRAWING THE ONLY SPECIAL BALL FROM n BALLS IN k TRIALS:**
Total # of outcomes = $^{n}C_{k}$ = $^{[1 + (n-1)]}C_{k}$ = $^{1}C_{0}{}^{(n-1)}C_{k}$ + $^{1}C_{1}{}^{(n-1)}C_{(k-1)}$ **, with term #1 denoting no special ball, and term #2 denoting the special ball**

***Total # of roundtable arrangements with k people = k!/k = (k-1)!**

*** SYSTEM RELIABILITY ANALYSIS:**

- P(fail)=p, P(success)=1-p
- For parallel config, 2^{n}-1 successes and 1 failure, P(fail)=p^{n}
- For series config, 2^{n}-1 failures and 1 success, P(success) = $(1-p)^{n}$
- For series connections, take intersection, and for parallel connections take union

*** PMF FOR SUM, DIFF, MAX, MIN OF 4-SIDED DICE:**

- Uniform PMF = $P_{XY}(x,y)$ = 1/16
- For each (x,y) point in the Cartesian coordinate diagram, calculate the diff/sum label or min/max label.
- Write down tables for Joint, Marginal and Conditional PMFs
- Headers are: x y $P_{XY}(x,y)$ x $P_{X}(x)$ y $P_{Y}(y)$ x y $P_{Y|X}(y,x)$. First 3 for joint, next 4 for marginal, the remaining for conditional
- For marginal, plot PMF on y-axis and RV value on x-axis.
- For joint, plot y on y-axis and x on x-axis

Facts for PMFs and RV Distributions

$0 \le P_X(x) \le 1$ ∀ x and **Sum over all x ∈ R_X ($P_X(x)$) = 1**

For any set A⊂R_X, P(X∈A) = $\sum_{x \in A} P_X(x)$

RVs X and Y are **independent if P(X=x, Y=y) = P(X=x) * P(Y=y), ∀ x,y** The first formula can be extended to n times.

P(Y=y|X=x) = P(Y=y), ∀ x,y if X & Y are independent

If $X_1,...,X_n$ are independent Bernoulli(p) RVs, then $X=X_1+X_2+...+X_n$ has Binomial(n,p) distribution, and **Pascal (1,p) = Geometric (p)**

For distributions using parameter p, **0<p<1**

If X is of Binomial (n, p = lambda/n), with fixed lambda > 0. Then, **for any k ∈ Z, $\lim_{n\to\inf} P_X(k)$ = ($e^{-lambda}$.$lambda^{k}$)/k!**

*** SPECIAL DISTRIBUTIONS:**

TYPE, PDF & E[X] AND VAR(X)

Uniform(a, b) **||** 1/(b-a) if a<x<b **||** (a+b)/2 and $(b-a)^2$/12

Exponential(lambda) **||** lambda . $e^{(-lambda . x)}$ **||** 1/lambda and 1/$(lambda)^2$

Normal/Gaussian, ie: N(0,1) **||** (1/sqrt(2 . pi)) . exp($-x^2$/2), ∀ x ∈ R **||** 0 and 1

Gamma (alpha, lambda) **||** ($lambda^{alpha}$. $x^{(alpha -1)}$. $e^{(-lambda . x)}$)/(alpha -1)! for x>0 **||** alpha/lambda and EX/lambda

CDF: $F_X(x)$ = P(X <= x) ∀ x ∈ R and **P(a < X <= b) = $F_X(b)$ - $F_X(a)$**

Counting Principles, n-nomial Expansions

Permutations of n distinct objs. take n w/ r groups of indistinct objs. = **(n!)/(n_1! ... n_r!)**

$^{n}P_{r}$ = n^{r} and $^{n}C_{r}$=$^{(n+r-1)}C_{r}$: for perms and combs where k objs are taken at a time

$(a+b)^{n}$ = Sum over k ($^{n}C_{k}a^{k}b^{(n-k)}$) where k=0,...,n

Binomial coeff. identity: $^{n}C_{k}$ = $^{(n-1)}C_{(k-1)}$ + $^{(n-1)}C_{k}$ where first term maps to A and second to A^{C}

$^{n}C_{m}$ = $^{n}C_{n-m}$

Sum over r ($^{n}C_{r}(-1)^{r}(1)^{(n+r)}$) is 0 where r=0,...n

Sum over r (($^{n}C_{r})^{2}$) = $^{2n}C_{n}$ where r=0,...,n

Sum over s ($^{s}C_{m}$) = ^(n+1)C~(m+1) where s=m,..,n

Hypergeometric expansion: $^{(n+m)}C_{r}$ = $^{n}C_{0}{}^{m}C_{r}$ + $^{n}C_{1}{}^{m}C_{(r-1)}$ + ... + $^{n}C_{r}{}^{m}C_{0}$ a CE and ME enumeration

n! = $(n/e)^{n}$ x root(2n x pi) - Stirling's approx. for n!

Trinomial expansion: $(a+b+c)^{n}$ = sum over i, j, k (C'$a^{i}b^{j}c^{k}$) where i, j, k=0,...n and i+j+k=n and C'=n!/(i!j!k!)

In n-nomial expansion $(a_1 + ... + a_r)^{n}$, the # of terms in the sum is $^{r}C_{n}$ = $^{(r+n-1)}C_{n}$ = $^{(r+n-1)}C_{(r-1)}$

Expectation, Variance, RV Functions

Expected value of X, ie: EX/E[X]/mu_X = sum over all x_k ∈ R_X (x_k . P(X = x_k)). It is linear

E[aX + b] = aE[X] + b, ∀ a,b ∈ R

E[X_1 + ... + X_n] = E[X_1] + ... + E[X_n]

If X is an RV and Y=g(X), then Y is also an RV.

Expectation, Variance, RV Functions (cont)

R_Y = {g(x) | x ∈ R_X} and P_Y(y) = sum over all x:g(x)=y (P_X(x))

E[g(X)] = sum over all x_k ∈ R_X (g(x_k) . P_X(x_k)) (LOTUS)

Var(X) = E[(X - mu_X)^2] = sum over all x_k ∈ R_X ((x_k - mu_X)^2.P_X(x_k))

SD(X)/$sigma_X$ = sqrt(Var(X))

Covariance = Cov(X,Y) = E[XY] - E[X].E[Y], which will be 0 if X & Y are independent

Var(X) = Cov (X,X) = E[X^2] - (E[X])^2

Var(aX + b) = a^2Var(X), and if X = X_1 + ... + X_n, then Var(X) = Var(X_1) + ... + Var(X_n)

Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)

Var(total up to X_n) = sum of all Var(X_i) if X_i is mutually independent for i = 1...n. Summing up over the same conditions for expected values holds true, regardless of independence or not

Correlation coefficient = Cov(X,Y)/(SD(X) . SD(Y)) - ranges between -1 and 1 (inclusive for both limits)

Z-standardized transformation: Z=(X - mu_X)/SD(X) - zero mean and unit variance