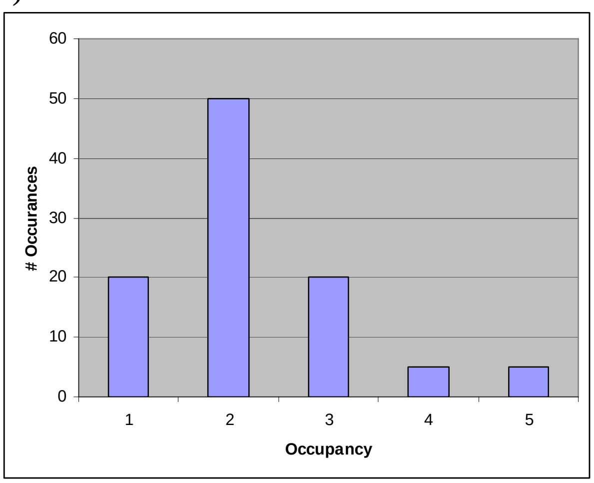
CE93 Midterm Solutions, October 30, 2007; Posted November 26, 2007
Prepared by Megan Smirti

Problem 1:
Occupancies of 100 vehicles driving in the HOV lane of the East Shore Freeway are observed. The results are summarized in the table below.

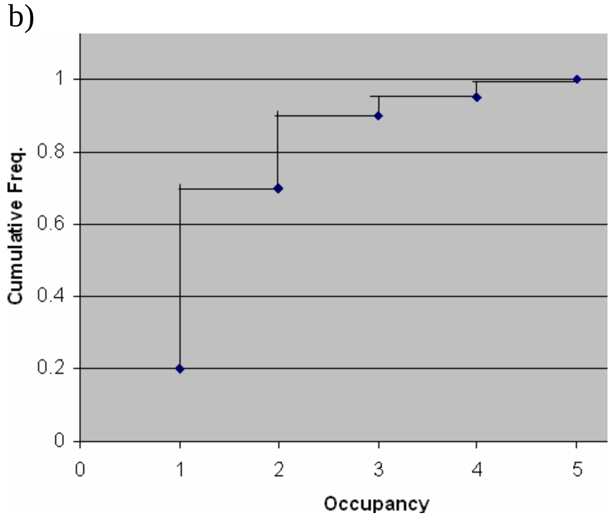
Occupancy (persons in vehicle)	Number of Vehicles
1	20
2	50
3	20
4	5
5	5

a) Draw a histogram for these data.
b) Draw a cumulative frequency diagram for these data.
c) Draw a box and whiskers plot for these data.
d) What is the median occupancy?
e) What is the variance of the occupancy?
f) Is the skewness of the occupancy positive or negative? Explain.
g) Does the variability in occupancy reflect aleatory or epistemic uncertainty? Explain

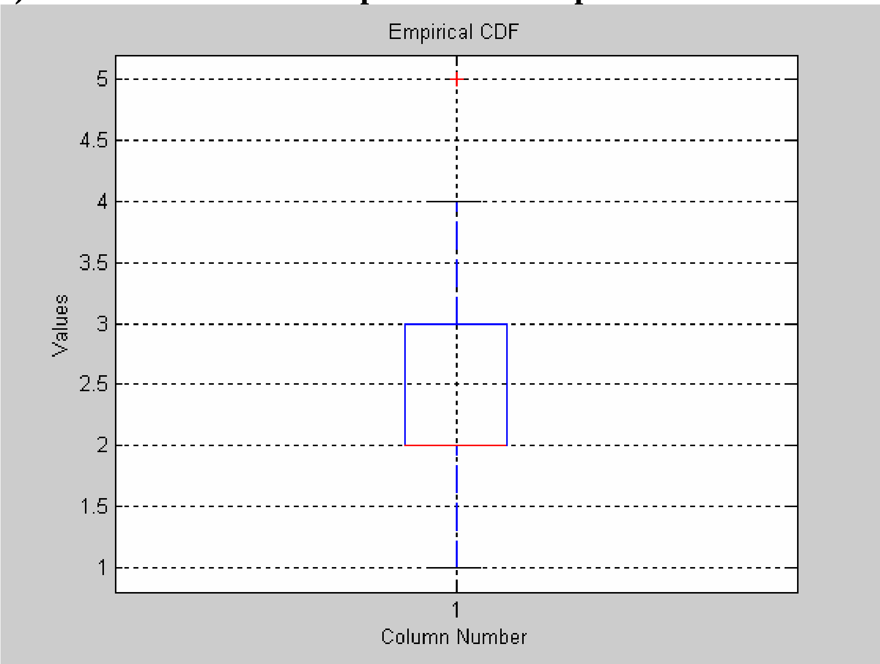
a)



b)



c) note – the outlier at 5 represents 5 data points



d) median is where $F_x(x) = .5 \rightarrow x_{.5} = 2$

e) $Var(X) = E(X^2) – E(X)^2$

E(X) =

1	*	0.2	+
2	*	0.5	+
3	*	0.2	+
4	*	0.05	+
5	*	0.05	=2.25

$E(X^2)$ =

1^2	*	0.2	+
2^2	*	0.5	+
3^2	*	0.2	+
4^2	*	0.05	+
5^2	*	0.05	=6.05

Var (X) = $6.05 - 2.25^2 = .988$

f) The skewness is positive. There are more possible values of occupancy to the right of the expected value. Also, the box and whiskers plot shows 5 outliers in the right tail of the distribution.

g) Aleatory uncertainty, or uncertainty associated with randomness, characterizes the variability of occupancy. This is because occupancy is a random variable, and as we collect occupancy data, we expect randomness. Epistemic uncertainty is uncertainty in knowledge – we have no reason to expect this uncertainty when collecting occupancy data.

CE93 Midterm Solutions, October 30, 2007; Posted November 26, 2007
Prepared by Megan Smirti

Problem 2:
At a construction project, the amount of material required in a day's construction is 200 units. The amount of material available on a given day is uniformly distributed between 0 and 500 units.

a) On a given day, it is known that the amount of material available is *at least* 150 units. What is the probability that not enough material is available?

P(x<200|x≥150) = $\frac{P(x<200 \cap x \geq 150)}{P(x \geq 150)} = \frac{50/500}{350/500}$ **= .143**

b) On a given day, it is known that the amount of material available is *less than* 150 units. What is the probability that not enough material is available?

P(x<200|x<150) = $\frac{P(x<200 \cap x<150)}{P(x<150)} = \frac{P(x<150)}{150/500}$ **= 1**

c) If not enough material is available, what is the probability that at least 150 units were available on that day?

P(x≥150|x<200) = $\frac{P(x \geq 150 \cap x<200)}{P(x<200)} = \frac{50/500}{200/500}$ **= .25**

Problem 3:
Suppose X and Y have the following joint distribution:

	X=0	X=1
Y=-1	0.1	0.2
Y=0	0.2	0.2
Y=1	0.1	0.2

a) What is the marginal distribution for X
P_x(x=0) = .1+.2+.1= .4
P_y(y=1) = .2+.2+.2=.6

b) What is the conditional distribution for Y with X=0?

$P_{y|x}$(y=-1| x=0) = $\frac{Px,y(y=-1 \cap x=0)}{Px(x=0)} = \frac{.1}{.4}$ **= .25**

$P_{y|x}$(y=0| x=0) = $\frac{Px,y(y=0 \cap x=0)}{Px(x=0)} = \frac{.2}{.4}$ **= .5**

$P_{y|x}$(y=1|x=0) = $\frac{Px,y(y=1 \cap x=0)}{Px(x=0)} = \frac{.1}{.4}$ **= .25**

c) What is the conditional distribution for X with Y=-1?

$P_{x|y}$(x=0|y=-1) = $\frac{Px,y(y=-1 \cap x=0)}{Px(y=-1)} = \frac{.1}{.3}$ **= .333**

$P_{y|x}$(x=1| y=-1) = $\frac{Px,y(y=-1 \cap x=1)}{Px(y=-1)} = \frac{.2}{.3}$ **= .667**

d) What is E(X)?

E(X) = 0*.4+1*.6 = .6

e) Are X and Y independent?

Test:
Does

$P_{x|y}$(x=0|y=-1) == P_x(x=0) ?

.333 = = .4 NO

So not independent, proven by counter example

f) What is the covariance of X and Y?

Cov(X,Y) = E(XY) – E(X)*E(Y)

E(X) = .6
E(Y) = -1*.3+0*.4+1*.3 = 0
So
E(X)*E(Y) = 0

E(XY)= Sum of these table entries:

0*-1*0.1	1*-1*0.2
0*0*0.2	1*0*0.2
0*1*0.1	1*1*0.2

Sum = 0

Cov(X,Y) = 0

g) Suppose we obtain 10 samples for X and Y drawn from this distribution, write a formula for the distribution of Z, the number of samples for X=Y=0.

$Z = 10 * P(x = 0 \cap y = 0) = 10 * .2 = 2$

Or

Z ~ Binomial

P = $P_{x,y}$ (0,0) = $P_{y|x}$(y=0| x=0) = .2

P_z (z) = $\binom{10}{z} p^z (1-p)^{10-z} = \binom{10}{z} .2^z (.8)^{10-z}$

Then maximize the function P_z (z) with respect to z, and get z=2

h) Suppose we keep drawing samples of X and Y from this distribution until we have 2 samples in which X=Y. What is the distribution for the of the number of samples we draw?

Z ~ Negative Binomial

P = $P_{x,y}$ (X=Y) = $P_{y|x}$(y=0| x=0) + $P_{y|x}$(y=1| x=1) = .4

P_z (z) = $\binom{z-1}{2-1} p^2 (1-p)^{z-2} = \binom{z-1}{1} .4^2 (.6)^{z-2}$

CE93 Midterm Solutions, October 30, 2007; Posted November 26, 2007
Prepared by Megan Smirti

Problem 4: The following are five samples of Random Walk code and six samples of Random Walk output. Match the correct code with the correct output diagrams. Provide a brief explanation for why the output matches the code. *Answer on this sheet, as show below:*

Code A
Output ____C______
Explanation:
The code presents a discrete distribution with three possible values. The output shows three discrete bars at each of the three possible values. There is no convergence to the normal distribution because n=1. Each observation in the histogram represents one random variable with a discrete distribution, and not a sum of many RVs.

Code B
Output ____F______
Explanation:
The code is a continuous exponential distribution with a large number of n. The output shows the continuous exponential distribution is converging to a normal distribution. In this example, n=100, which means each observation on the histogram is the sum of 100 RV with an exponential distribution. The central tendency is around 500, which is consistent with our expected value of the exponential distribution.

Code C
Output _____A_____
Explanation:
This is a continuous exponential distribution with n=10000, which is larger than code B. The output shows the expectation is around $5*10^4$, which is expected given our parameter for expected value and our number of n.

Code D
Output _____B_____
Explanation:
This is a discrete distribution converging to the normal distribution because n=100. The central tendency is around 125 which is consistent with the potential three values which can occur with certain probabilities.

Code E
Output _____E_____
Explanation:
The code is a continuous exponential distribution with n=1. The histogram looks exactly like an exponential distribution. In the code n=1, and therefore the exponential distribution is not converging to the normal distribution because each observation represents one exponentially distributed RV and not the sum of many exponentially distributed RVs.