Stat 134 MIDTERM (Sec 3, Spring 2009) J. Pitman.

Name and SID number: ______________________________

Please circle final answers. Use additional space to provide explanation.

1. A deck of 4 cards contains one card numbered 1, two cards numbered 2, and one card numbered 3. The deck is shuffled thoroughly. Let D be the difference between the numbers on the top two cards, ignoring the sign.
   a) Display the distribution of D in a suitable table.

d	0	1	2
$P(D=d)$	$2/12$	$8/12$	$2/12$

   b) Find the variance of D.

$$E(D-1)^2 = \frac{4}{12}\cdot 1^2 = \frac{1}{3}$$

2. Consider independent rolls of a fair six-sided die.
   a) Sketch the probability histogram of the number of sixes in 180 rolls.

$$\mu = np = 180 \times \frac{1}{6} = 30$$

$$\sigma = \sqrt{npq} = \sqrt{180 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 5$$

   b) Find approximately the probability of ~~180~~ 35 or fewer sixes in 180 rolls.

$$\Phi\left(\frac{35\frac{1}{2} - 30}{5}\right) = \Phi(1.1) \simeq 0.86$$

3. A manufacturing process produces sheets of glass which contain bubbles, with on average one bubble per five square feet. Assume a homogeneous Poisson random scatter of bubbles.
   a) Window panes of one square foot each are considered of acceptable quality only of they contain no bubbles. What fraction of window panes are of acceptable quality?

$$e^{-1/5}$$

$\lambda = 1/5$ per sq ft
area $= 1$ sq ft
$\mu = \lambda \times \text{area} = 1/5$

   b) Suppose each bubble is large with probability p and small with probability $1-p$, independently from bubble to bubble, and the acceptance policy is revised so panes are accepted if they contain no large bubbles. What will then be the fraction of panes of acceptable quality?

$e^{-(1-p)/5}$ Thinning property.

4. A spam filter works by first classifying incoming email messages into one of two exclusive categories C_i which are found empirically to have long-run relative frequencies p_i for $i = 1, 2$. Of messages in class C_i, the proportion of messages that are spam is found to be s_i.
a) What is the long-run frequency of spam messages in this email stream?

$$p_1 s_1 + p_2 s_2$$

b) The spam filter stops all email in category C_1 and lets through all email in category C_2. Of all the email messages that pass through the filter, what proportion is spam?

$$s_2$$

5. Suppose that k balls are thrown independently and uniformly at random into $n \geq 2$ boxes. Let X be the number of empty boxes.
a) Find a formula for $E(X)$.

$$n\left(1-\frac{1}{n}\right)^k$$

because $X = X_1 + \cdots + X_n$, $P(X_i = 1) = \left(1-\frac{1}{n}\right)^k$

b) Find a formula for $E(X^2)$.

X_i = indic box i is $\emptyset$.

$$EX^2 = n\,EX_1^2 + n(n-1)\,E(X_1X_2)$$
$$= n\left(1-\frac{1}{n}\right)^k + n(n-1)\left(1-2/n\right)^k$$

6. A pair of positive integer valued random variables X and Y has

$$P(X = n, Y = m) = cr^{n+m} \text{ for } n, m = 1, 2, \ldots \text{ for some } c > 0 \text{ and } 0 < r < 1.$$

a) What is the distribution of X?

Geometric (p) for $p = 1 - r$

b) Give a formula for $E[(X - Y)^2)]$ with no unsimplified sums.

$$E\left((X-Y)^2\right) = \mathrm{Var}(X-Y) = 2\,\mathrm{Var}(X)$$
$$= \frac{2q}{p^2} = \frac{2r}{(1-r)^2}$$