STT 200
Exam 3 Prep KEY

About Exam 3:

Usual assigned seating (posted on website).
No electronics in use or in view by anyone.
No notes or books in use or in view by anyone.
No writing on hands, etc.
Fill out cover page when you get the exam.
Do not open the exam until the signal to begin.
Know your seat number and section number!
On the white exam paper give your reasoning/calculations.
Points may be withdrawn for answers without substantiation.
Keep your eyes on your own work.
You may be asked to move.
Do not put your work in view of others.
Exam 3 is in B106 Wells on Wednesday 3-26-08.
A key to this prep will be posted late Thursday 3-20-08.
Chapter 15, 16, and Binomial & Poisson are covered.
I will not examine the formula for P(X = x) pg. 388.
Standard scores (z-scores) and z-table use are covered.

1. Classical probability. List all 8 outcomes for three tosses of a coin. From this list, just by counting favorable cases, determine

a. P(total of 2 heads)
ans. P(HHT or HTH or THH) = 3/8

b. The probability distribution of random variable X = number of heads in the three tosses.

x	0	1	2	3
p(x)	**1/8**	**3/8**	**3/8**	**1/8**

c. P(X < 2) **= p(0) + p(1) = 4/8**

2. Classical and Binomial Model. Note: I've changed from {0.3, 0.7} to {0.2, 0.8}. A particular biased coin has probability 0.2 of H and probability 0.8 of T each time it is thrown, tosses being independent. Give

a. P(HHH), P(HHT), P(HTH), P(HTT)
$.2^3$ $.2^2.8$ $.2^2.8$ $.2\,.8^2$

P(THH), P(THT), P(TTH), P(TTT)
$.2^2.8$ $.2\,.8^2$ $.2\,.8^2$ $.8^3$

b. The distribution of random variable X = number of heads produced by this biased coin in three throws (by gathering cases in (a)).

x	**0**	**1**	**2**	**3**
p(x)	**$.8^3$**	**$3\,.2\,.8^2$**	**$3\,.2^2.8$**	**$.2^3$ (total 1)**

c. Consulting the formula for the binomial probability model on page 384, evaluate $P(X = 2) = {}_3C_2\, 0.2^2\, 0.8^{3-2}$ and compare with p(2) from part (b).
$${}_3C_2\, 0.2^2\, 0.8^{3-2} = \frac{3!}{2!\,(3-2)!} 0.2^2\, 0.8^{3-2} = 3\; 0.2^2\; 0.8$$

3. Binomial and Normal. According to the panel on page 385, the mean of a binomial distribution is np and the standard deviation is $\sqrt{npq}$. For n = 100 tosses of the biased coin (2) the possible number of heads varies between 0 and 100. **The approximate distribution will be a bell curve with mean np and standard deviation $\sqrt{npq}$ where p = 0.2 and q = 1 - p = 0.8.**

a. Evaluate these quantities exactly and sketch the bell curve. Special Binomial calculations of μ, σ apply.
np = 100 0.2 = 20 (on avg, expect 20 heads in 100 tries)
sd = $\sqrt{npq} = \sqrt{100\;0.2\;0.8} = \sqrt{16} = 4$
(bell curve with mean μ = 20 and sd σ = 4)

b. From (a) determine (approximately) P(X > 28) by calculating the standard score z of x = 28 and appealing to the z-table to determine P(Z > z).
$$z = \frac{28-\mu}{\sigma} = \frac{28-20}{4} = 2 \qquad P(Z > 2) \sim 0.025$$

Note: P(X > 28) = p(29) + p(30) ... + p(100). The bell approximation of p(29) is roughly the area of the bell curve between limits [28.5, 29.5] (think as in calcu-

lus). So a better normal approximation of P(X > 28) instead uses "continuity corrected" standard score

$z = \frac{28.5 - \mu}{\sigma} = \frac{28.5 - 20}{4} = 2.125$ P(Z > 2.125) ~ 0.017

As you can see, continuity correction can make a difference in the normal approximation of P(X > 28). We will occasionally use continuity correction in homework but not on the exams.

4. Poisson and Normal. We expect on average around 16 customers in a given region of the store. That region can comfortably hold 24 customers. Suppose the random variable X = number of customers in the region is Poisson distributed.

According to page 388 the standard deviation of the Poisson distribution is $\sigma = \sqrt{\mu}$. Moreover, **the Poisson distribution is approximately normal for $\mu \geq 3$ (a rule of thumb)**.

a. Determine the mean and sd of the distribution of the number of customers in the specified region of the store and sketch the approximate distribution.

mean = 16 (> 3) sd = $\sqrt{\text{mean}} = \sqrt{16} = 4$
Bell curve with mean 16 and sd 4.

b. Calculate the standard score z for 24 (the comfort limit) and approximate P(X > 24) by P(Z > z) using the z-table. Is there much chance of exceeding the comfort limit of the region?

$P(X > 24) \sim P(Z > \frac{24 - 16}{4}) = P(Z > 2) \sim 0.025$

Note: Continuity correction can help here too. It would be more accurate to use P(X > 24) ~ P(Z > z) for $z = \frac{24.5 - 16}{4}$ = 2.125, obtaining a more accurate approximation P(X > 24) ~ P(Z > 2.125) ~ 0.017 from the z-table.

Note on using z-table: It is convenient to do it this way,
P(Z > 2.125) = P(Z < -2.125) (symmetry of Z around 0)
~ 0.017
(directly from the z-table entry for z = -2.12 or z = -2.13)

5. Means, Var, sd of r.v. Random variables X, Y have

	X	Y
mean	12	7
sd	10	9

a. Determine E(3X+8Y-2) (does not require X, Y independent)
ans. 3 EX + 8 EY - 2 = 3(12)+8(7)-2 = 90

b. If X, Y are **independent (important)**, what are
Var(3X+8Y-2) **= 9 Var X + 64 Var Y = 9(100) + 64(81)**
Var(2X-4Y) **= 4 Var X + 16 Var Y = 400 + 16 81 = 1696**
(add variances because Var(-4Y) = 16 Var Y)

c. If r.v. X, Y are independent E(XY) = (E X)(E Y). Evaluate it.
E XY = EX EY = 12(7) = 84

6. Sums of many independent ventures ~ normal. Examining our month by month e-business profits it seems we average \$2200 profit per month with a sd of \$1500. It seems also that our monthly profits are essentially independent of one another (nothing is learned about an unseen month by consulting other months).

a. Let r.v. X denote the total profit for the next 12 months. Determine

E X **= 12 2200 = \$26,400**

Var X = **12 1500^2 = 27,000,000**
(remember, variance of a total of independent r.v. is the sum of the individual variances)

sd X **= $\sqrt{27,000,000}$ = 5,196.15**
So we expect to earn \$26,400 on average with an sd of \$5,196.

b. Sketch the approximate distribution of X using the fact that **a sum of many independent r.v. tends to have an approximately normal distribution** (look ahead to pg. 415 if you wish to see a formal statement, but for now we just use it as a rule of thumb).

Bell curve with mean 26,400 and sd 5,196.15.

c. From (b) use the standard score z of x = 25,000 to approximate P(X < 25,000) by P(Z < z) using the z-table. Can we be fairly comfortable planning to earn at least \$25,000?

P(X < 25000) ~ **P(Z < $\frac{25\,000-26\,400}{5196.15}$) = P(Z < -0.27) ~ 0.39.**

No, there is around 39% chance we won't!

7. OIL. We are given

P(OIL) = 0.4
P(+ | OIL) = 0.7
P(+ | OIL^C) = 0.2
cost to test = 100
cost to drill = 600
gross from drilling if OIL is present = 1400

a. Make a compete tree diagram with all branch probabilities and all endpoint probabilities. To the right of each endpoint probability place the calculation of the net return from the policy "test, but only drill if the test is positive" for that contingency.

event	pr	calculation of net
OIL+	**0.4 0.7 = 0.28**	**-100 - 600 + 1400 = 700**
OIL-	**0.4 0.3 = 0.12**	**-100 - 0 + 0 = -100**
OIL^C+	**0.6 0.2 = 0.12**	**-100 - 600 + 0 = -700**
OIL^C-	**0.6 0.8 = 0.48**	**-100 - 0 + 0 = -100**

b. Calculate E X where X is the net return from the policy of (a).

E NET **= .28 700 - .12 100 - .12 700 - .48 100**
= 196 - 12 - 84 - 48 = 196 - 144 = 52

c. Calculate the risk σ = sd X for the policy of (a).

Var NET = E NET^2 − (*E* NET)^2
= 202,000 - 2704 = 199,296

where

E NET^2 = .28 700^2 + .12 100^2 + .12 700^2 + .48 100^2
= 202,000

So then

σ **= sd NET = $\sqrt{199,296}$ = 446.426**

d. Calculate the expected net return and risk = sd of net return for the policy "just drill." The simplest model (i.e. just OIL and OIL^C, without the "test" sub-branches) suffices for this calculation.

event	pr	calculation of NET
OIL	0.4	**-600 + 1400 = 800**
OIL^C	0.6	**-600 + 0 = -600**

E NET = 0.4 800 - 0.6 600 = 320 - 360 = -40

Var NET = $\mathbf{E\,NET^2 - (E\,NET)^2}$ **= 472,000 - 1600 = 470,400**

where

$$\mathbf{E\,NET^2 = .4\;800^2 + .6\;600^2 = 472{,}000}$$

So then

$$\mathbf{sd\;NET = \sqrt{470,400} = 685.9}$$

e. Is it clear to you which of the two policies is the preferred one? **Policy "just drill" has negative expectation and larger risk (i.e. sd). So it seems to be the worst of the two policies. But the policy "test, then drill if test is positive" is also extremely risky!**

8. Drawing Balls. A box has colored balls {5 R, 8 G, 2 Y}. Draws will be made without replacement with equal probability on the balls then remaining.

a. P(R2) by the principle "order of the deal does not matter" is

$$\mathbf{= P(R1) = \frac{5}{5+8+2} = \frac{1}{3}}$$

(draw 2 is like draw 1)

b. Calculate

$P(R1^C\,R2)$ = $\mathbf{P(R1^C)\;P(R2 \mid R1^C)}$

$$\mathbf{= \frac{10}{15}\;\frac{5}{14}}$$

$P(R1\,R2)$ = $\mathbf{P(R1)\;P(R2 \mid R1)}$

$$\mathbf{= \frac{5}{15}\;\frac{4}{14}}$$

c. From (b) determine P(R2) and compare with (a).

$P(R2)$ = $\mathbf{P(R1^C\,R2) + P(R1\,R2)}$ **(the only ways to get R2)**

$$\mathbf{= \frac{10}{15}\;\frac{5}{14} + \frac{5}{15}\;\frac{4}{14} = \frac{5}{15}\left(\frac{10}{14} + \frac{4}{14}\right) = \frac{5}{15} = \frac{1}{3}}$$

d. Calculate P(R2 or Y2) = P(R2 $\cup$ Y2).
P(R2 or Y2) **= P(R2) + P(Y2) - P(R2 and Y2) (addition rule)**
$= \frac{5}{15} + \frac{2}{15} - 0 = \frac{7}{15}$ (like draw 1)

9. Poisson Cookies. A batch of 400 cookies has 1600 chocolate chips stirred in.

a. What are the mean and sd of the number of chips in a given cookie? **We average $\frac{1600}{400} = 4$ chips per cookie. So $\mu = 4$. For the Poisson distribution (on counts of rare events where applicable) we have sd $\sigma = \sqrt{\mu} = \sqrt{4} = 2$.**

b. Since the mean (a) is at least 3 we have (by the rule of thumb) the normal approximation for the distribution of the number of chips in a given cookie. Sketch it, indicating mean and sd.
Bell curve with mean 4 and sd 2.

c. What is the approximate probability that a given cookie has fewer than 2 chips? **Score x = 2 is one sd below the mean of 4. That is, its standard score is z = $\frac{2-\mu}{\sigma} = \frac{2-4}{2} = -1$.**

Since P(Z < -1) = 0.16 we reckon (by using the normal approximation in this straightforward way) there is around 16% chance a given cookie would contain fewer than 2 chips.

NOTE. Actually, a far better approximation is to instead use the "continuity corrected" z score $\frac{1.5-4}{2}$ = -1.25. This is because "fewer than 2" means "0 or 1 chip" which is better approximated under the bell curve by the area left of 1.5 than the area left of 2. This gives P(fewer than 2 chips in a given cookie) ~ P(Z < -1.25) ~ .106 (not 0.16).

We will NOT use the continuity correction on exams (even though it is often far more accurate!) but will sometimes use it in homework since it is the best way to

go.

d. From (c) around how many of the 400 cookies will have fewer that 2 chips? **Since there is around 0.106 probability that any given cookie has fewer than 2 chips we expect 400 0.106 = 42.4 cookies of that type in a batch of 400.**

On an exam you would NOT use the continuity correction. But keep in mind that doing so gives the far less accurate answer 400 0.16 = 64 cookies of this type expected in a batch of 400.

Using the normal approximation for discrete distributions can introduce these inaccuracies. In the case of Poisson or Binomial we have recourse to the continuity correction which goes a long way towards fixing the problem. BUT WE'LL NOT USE THE CONTINUITY CORRECTION ON EXAMS.

10. Probability Rules. P(rain today) = 0.7, P(rain tomorrow) = 0.8, P(rain tomorrow | rain today) = 0.9. Determine

a. P(rain both days)
= P(today) P(tomorrow | today) = 0.7 0.9 = 0.63

b. Are these independent events? Why?
They are NOT independent since P(tomorrow) = 0.8 is not the same as P(tomorrow | today) = 0.7. In other words, knowing it rains today will change the probability of rain tomorrow. Another way to recognize this lack of independence is that from the above P(rain both days) = 0.63 which is NOT the same as P(today) P(tomorrow) = 0.7 0.8 = 0.56 (the wrong answer, precisely because the events are NOT independent and you are therefore not entitled to simply multiply by the unconditional probability of rain tomorrow).

c. P(rain today $\bigcup$ rain tomorrow), i.e. it rains today **or** tomorrow and possibly both days.

P(rain today $\bigcup$ rain tomorrow)
= P(today) + P(tomorrow) - P(both days)
= 0.7 + 0.8 - 0.63 = 0.87

d. P(today and tomorrow) **= 0.63 (see above)**
P(today and not tomorrow) **= P(today) - P(both)**
= 0.7 - 0.63 = 0.07

P(not today, and tomorrow)
= P(tomorrow) - P(both) **= 0.8 - 0.63 = 0.17**

P(neither day) **= 1 - (0.63 + 0.07 + 0.17) = 0.13**
(or, just 1 - P(today or tomorrow) = 1 - 0.87 = 0.13)

Make the Venn diagram for these probabilities.

11. Probability Rules and Independence. Assume a job has two tasks which must both be completed if the job is to get done. These events are independent with respective probabilities 0.8, 0.6.

a. What is the probability both tasks are completed (hence the job gets done)? **By independence,**
P(task 1) P(task 2 | task 1) = P(task 1) P(task 2)
= 0.8 0.6 = 0.48

b. What is the probability at least one task is completed?
P(task 1 or task 2) **= P(task 1) + P(task 2) - P(both)**
= 0.8 + 0.6 - 0.48 (see above)
= 1.4 - 0.48 = 0.92

12. More probability. You will choose box A or box B (below) with respective probabilities 0.3 and 0.7. Following that, you will select two balls without replacement from the selected box. Determine P(R1 Y2).

box A: {4R 6Y}
box B: {9R 8G 4Y}

P(R1 Y2)
= P(boxA R1 Y2) + P(boxB R1 Y2)
= P(boxA) P(R1 Y2 | boxA) + P(boxB) P(R1 Y2 |box B)
$$= 0.3 \frac{4}{10} \frac{6}{9} + 0.7 \frac{9}{21} \frac{4}{20}$$