This week we finish random variables, expectation, variance and standard deviation. We also begin "tests of statistical hypotheses" on Wednesday. Read "Testing Hypotheses about Proportions" in your textbook before Wednesday 12-2-09.

1-4. Three Doors Problem (Monte Hall Problem). The host of a TV program presents three "doors" only one of which conceals a major prize. A contestant (you) is invited to choose one of these three doors. You choose. The host then opens a different door behind which the major prize is not found. You are invited to either KEEP the door you originally chose or SWITCH to the remaining door (not the one you originally chose and not the one opened). Should you switch? Assume that whatever is behind each of the three doors does not change during the course of play and will be revealed if any particular door is opened by the host. As for whether that is the case for the TV game you may get some clues by listening carefully to what is actually said by the host in the course of play. When we speak of "the prize" it is the major prize (lesser prizes may be behind some doors and other offers made but we ignore them).

1. There is uncertainty but no probability in the problem as stated. To illustrate some possible behaviors of host and contestant, suppose that you intend to choose door 1 (of doors 1, 2, 3), the host has placed the prize behind door 3, and the host plans to open door 2 if you choose door 3, but open door 1 if you choose door 2.

a. List all possible outcomes depending on whether you keep or switch.

prize located	contestant (initially) chooses	host reveals	switch?

b. From (a) determine the probability that you will win the major prize *if you flip a coin* to decide whether to keep or switch. Note that these probabilities flow from the contestant's use of a coin.

2. Instead of (1) assume that *the contestant* (you) *initially chooses one of the three doors by a random equal probability draw* from {1, 2, 3}.

a. Complete the list of **all twelve possible choices** of a triplet consisting of (a door behind which the major prize may be, the door selected by the contestant, and the door revealed by the host). Remember, as described in the initial problem statement we are dealing only with the case in which the host never opens the prize door until the end.

prize located	contestant (initially) chooses	host reveals
1	1	2
1	1	3
1	2	3
1	3	2
2	2	1

etc.

This is not a probability model since we do not have probabilities on all of the twelve possible outcomes. The only probabilities we have are the probabilities 1/3 of the contestant choosing each particular door.

b. The contestant may adopt a policy of *selecting a door at random as in (a) and always keeping it (i.e refusing to switch)*. Call this the KEEP policy. Is it clear to you that P(KEEP wins) = P(contestant initially chooses the door behind which the prize is located)? What then, must be the value of P(KEEP wins)?

c. A probability model for (b) can be built by first agreeing upon the premise that since the contestant chooses at random we may as well suppose that the prize is behind door 1. The sample space of the experiment is then {1, 2, 3} and the contestant wins if 1 is selected. Verify that your answer to (b) agrees with the value of P(KEEP wins) calculated from this model.

d. Refer to (c). Consider the policy SWITCH in which, having chosen the initial door at random, the contestant *always switches* when offered the chance to do so. Observe that in model (c) SWITCH will always win when KEEP loses and visa-versa. If so, what is P(SWITCH wins)?

e. Suppose the prize is $10,000 (otherwise $0). Determine the *probability distribution* of the random variable X = amount won by the KEEP policy. From the distribution of X determine E X, Var X, SD X.

f. As (e) but for Y = amount won by the SWITCH policy. From the distribution of Y determine E Y, VAR Y, SD Y.

3. A random variable W has E W = 41.7, SD X = 17.2.

a. Since the distribution of W has not been given we cannot use it to determine the value of E (3W + 2), VAR (3W + 2), SD (3W + 2) directly from the distribution. We can however obtain these using the rules and the information already given. Do so.

b. Since the distribution of (3W + 2) is not known we cannot sketch it. If however the distribution of W (and therefore of 3W + 2) is *approximately normal* we can. Do so, labeling the mean and sd of (3W + 2) in the appropriate places of your sketch.

c. Even if the distribution of W is **NOT** known to be approximately normal *the distribution of the sum of a large number n of independent random variables, each having the distribution of* W, WILL BE approximately normal with mean = n (E W) and sd = ($\sqrt{n}$ sd W). Provide a sketch of the approximate distribution of $W_1 + \; \ldots \; + W_{100}$ where the 100 summands are *independent* random variables each having the distribution of W. Label the mean and sd at appropriate places in your sketch.

4. John is at a casino where he will play 100 rounds of game A and 400 rounds of game B. Any single play of A has expected net return E X = -$.035 with SD X = $3.65. Any single play of B has expected net return E Y = -$.024 with SD Y = $4.11. The large standard deviations are due to some possible returns being very positive and large although they may have extremely small probability.

a. Assume all 500 plays are independent and that the normal approximation will apply to John's total winnings. Sketch the normal approximation of the distribution of John's winnings? Use

$E(X_1 + \ldots + X_{100} + Y_1 + \ldots + Y_{400}) = 100 \text{ E } X + 400 \text{ E } Y$

independence not required between plays

$VAR(X_1 + \ldots + X_{100} + Y_1 + \ldots + Y_{400}) = 100 \text{ VAR } X + 400 \text{ VAR } Y$

independence between plays is used

$SD(X_1 + \ldots + X_{100} + Y_1 + \ldots + Y_{400}) = \sqrt{100 \text{ VAR } X + 400 \text{ VAR } Y} =$

b. Instead of spending so much time on 500 plays John decides to put 100 times as much money on one play of A and 400 times as much money on one play of B. So his total winnings can now be expressed as 100 X + 400 Y. Assuming X, Y are independent give

$E(100 X + 400 Y) = 100 \text{ E } X + 400 \text{ E } Y =$

$VAR(100 X + 400 Y) = 100^2 \text{ Var } X + 400^2 \text{ Var } Y =$

$SD(100 X + 400 Y) = \sqrt{\text{VAR}(100 X + 400 Y)} =$

Compare with the values expectation and variance and standard deviation of aggregate return $(X_1 + \ldots + X_{100} + Y_1 + \ldots + Y_{400})$ in (a).

c. Using many plays as in (a) John is essentially buying entertainment and a very small chance of either emerging as a winner at the end of play or "losing his shirt."

Use (a) and the normal table to approximate:

P(John's total winnings are positive) ≈

P(John's total winnings are below -$200) ≈

5. Not related to #4. Suppose that in one play X you win $1,000 with probability .000965 on an investment of $1, otherwise losing your dollar. So NET return X follows the distribution

value x	**$999**	**-$1**
p(x)	**.000965**	**1-.000965**

a. Determine E X, VAR X, SD X.

b. How much could John win betting $100 in one round (i.e. 400 X)? What is the probability he does so? The strategy of lumping $400 on one play of net return X can reap enormous returns but with small probability.

c. Determine E (400 X), Var (400 X), SD (400 X).

6. OIL example. Suppose

P(OIL) = 0.28 **P(+ | OIL) = 0.88** **P(+ | OIL^C) = 0.17**

cost to test = 80 **cost to drill = 900** **gross return from OIL = 3000**

a. Fill out a complete tree diagram with all branches and endpoints properly labeled and with all their respective probabilities as well.

b. Use the endpoint probabilities of your tree to complete a properly labeled Venn diagram with all four regions properly labeled and with their probabilities in place.

c. On the tree diagram label also all values of X = net return from the policy "test, but drill only if the test is positive."

d. Using (c) calculate E (net return from policy of (c)).

e. Compare (d) with E(net return from policy "just drill").