Stat 110 Midterm

Prof. Joe Blitzstein

October 12, 2011

This exam is closed book and closed notes, except for two standard-sized sheets of paper (8.5" by 11") which can have notes on both sides. No copying, cheating, collaboration, calculators, computers, or cell phones are allowed. Show your work. Answers should be exact unless an approximation is asked for. All parts will be weighted equally within each problem. The last page contains a table of important distributions, and the page before that can be used for scratch work or for extra space. If you want any work done on the extra page or on backs of pages to be graded, mention where to look in big letters with a box around them, on the page with the question. Good luck (an appropriate expression for this course)!

Name Ravi Dayabhai

Harvard ID 40 9066394

Prior probability, $P(A)$, that new treatment is better: $P(A) = \frac{2}{3}$. B is event that treatment was effective for 15 out of 20 random patients.

$T \sim \text{Bin}(20, 60\%)$, $B \equiv T = 15$.

1. A new treatment for a disease is being tested, to see whether it is better than the standard treatment. The existing treatment is effective on 50% of patients. It is believed initially that there is a 2/3 chance that the new treatment is effective on 60% of patients, and a 1/3 chance that the new treatment is effective on 50% of patients. In a pilot study, the new treatment is given to 20 random patients, and is effective for 15 of them.

(a) Given this information, what is the probability that the new treatment is better than the standard treatment (as an unsimplified number)?

$$p = P(A|B) = \frac{P(B|A)P(A)}{P(B)} =$$

(assume effectiveness is independent by patients)

$$\boxed{\frac{\frac{2}{3}\binom{20}{15}(.6)^{15}(.4)^{5}}{\frac{2}{3}\binom{20}{15}(.6)^{15}(.4)^{5} + \frac{1}{3}\binom{20}{15}(.5)^{20}}}$$

(b) A second study is done later, giving the new treatment to 20 new random patients. Given the results of the first study, what is the PMF for how many of the new patients the new treatment is effective on? (Don't simplify; letting p be the answer to (a), your answer can be left in terms of p.)

Let X be r.v. counting how many of patients the new treatment was effective on of the 20 new random patients. Knowing results from pilot, ~~we have $X \sim \text{Bin}(20, r)$ where r is updated probability of effectiveness~~ : not binomial — a mixture of binomials!

~~$r = p(.6) + (1-p)(.5)$, so $P(X=x) = \binom{20}{x} r^x (1-r)$~~.

The patient outcomes are <u>conditionally</u> independent on effectiveness, but not unconditionally independent!

INDEPENDENCE $\not\Leftrightarrow$ CONDITIONAL INDEPENDENCE.

NOT 3 teams of 120 people!

Correct reasoning, not paying attention!

X 2. A group of 360 people are going to be split into 120 teams of 3 (where the order of teams and the order within a team don't matter).

(a) How many ways are there to do this (simplified, in terms of factorials)?

$$\cancel{\frac{\binom{360}{120}\binom{240}{120}}{3!} = \frac{\frac{360!}{120!\,240!}\cdot\frac{240!}{120!\,120!}}{3!} = \boxed{\frac{360!}{3!\,(120!)^3}}}$$

$$\boxed{\frac{360!}{(3!)^{120}(120!)}} = \frac{\binom{360}{3}\binom{357}{3}\cdots\binom{3}{3}}{120!}$$

X (b) The 360 people consist of 180 married couples. A random split into teams of 3 is chosen, with all possible splits equally likely. Find the expected number of teams containing married couples. (You can leave your answer in terms of binomial coefficients and a product of a few terms, but you should not have summations or complicated expressions in your final answer.)

Let I_j be indicator r.v. for $i \in \{1,2,3\}$ where I_j indicates whether team j contains at least 1 married couple. $X = I_1 + I_2 + I_3$, and by LINEARITY, $E(X) = E(I_1) + E(I_2) + E(I_3)$. By FUND. BRIDGE & SYMMETRY, this is the same as $E(X) = 3 \cdot P(\text{at least 1 couple on Team 1})$.

$$\underbrace{1 - P(\text{no couples on Team 1})}_{\frac{\binom{180}{120}2^{120}}{\binom{360}{120}}}$$

Correct reasoning, not paying attention!

$$\boxed{3\left(1 - \frac{\binom{180}{120}2^{120}}{\binom{360}{120}}\right)}$$

3. Let $\boxed{X \sim \text{Bin}(100, 0.9)}$. For each of the following parts, construct an example showing that it is possible, or explain clearly why it is impossible. In this problem, Y is a random variable on the same probability space as X; note that X and Y are not necessarily independent.

(a) Is it possible to have $Y \sim \text{Pois}(0.01)$ with $P(X \geq Y) = 1$?

No because the support for Y is unbounded – although it is unlikely that $Y > X$, it could occur, so $P(X \geq Y) \neq 1$. X is bounded at 100, and despite the $P(Y > 100)$ being small, it is non-zero!

$\frac{1}{2}$ credit.

(b) Is it possible to have a $Y \sim \text{Bin}(100, 0.5)$ with $P(X \geq Y) = 1$?

If X and Y are independent, this is impossible because each product of $P(X=x)\,P(Y=y) = P(X,Y) > 0$ for all $x, y \in \{1, .., 100\}$.
If dependent, this is possible... can't come up with example.
See "chicken and egg" homework problem!
e.g. X counts # eggs hatching (prob. of hatch $= 0.9$) and
Y counts # of chicks surviving (prob. of surv.|hatch $= .5$)

$\frac{1}{2}$ credit (c) Is it possible to have $\boxed{Y \sim \text{Bin}(100, 0.5)}$ with $P(X \leq Y) = 1$?

If X and Y are independent, this is impossible (same reason as above).
If dependent, not sure.

$P(X \leq Y) \Rightarrow P(0 \leq Y - X) = 1 \Rightarrow E(Y-X) \geq 0 \Rightarrow E(Y) - E(X) =$

$50 - 90 \geq 0$ X impossible.

4. Athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with $X_1, X_2, \ldots$ i.i.d. with $\ln(X_j) \sim \mathcal{N}(\mu, \sigma^2)$. We say that the jth jumper is "best in recent memory" if he or she jumps higher than the previous 2 jumpers (for $j \geq 3$; the first 2 jumpers don't qualify).

(a) Find the expected number of best in recent memory jumpers among the 3rd through nth jumpers (simplify).

almost!

(SEE NEXT PAGE).

$\boxed{E(Y) = (n-2)\frac{1}{3}}$ where Y counts number of best-in-recent-memory jumpers.

(b) Let A_j be the event that the jth jumper is the best in recent memory. Find $P(A_3 \cap A_4)$, $P(A_3)$, and $P(A_4)$ (simplify). Are A_3 and A_4 independent?

$\boxed{P(A_3) = E(I_3) = \frac{1}{3}}$

$\frac{1}{3} = P(A_4) = P(A_4 | A_3) P(A_3) + \boxed{P(A_4 | A_3^c) P(A_3^c)}$

$= \cancel{\frac{1}{3} \cdot \frac{1}{3}} + \frac{1}{4} \cdot \frac{2}{3}$

$= \frac{2}{4!} = \frac{1}{12}$

correct but for wrong $P(A_3, A_4)$ and $P(A_4)$.

$\boxed{\text{not independent b/c } P(A_3, A_4) = P(A_4 | A_3) P(A_3) = \frac{1}{9} \neq \frac{1}{3} \cdot \frac{1}{4} = P(A_3) P(A_4)}$

(c) Find the variance of X_j (you can leave your answer in terms of integrals).

Let $X = X_j$, $Y = \ln(X_j) = \ln(X)$. Then $\text{Var}(X) = \text{Var}(e^Y) = E(e^{2Y}) - E(e^Y)^2 =$ right idea...

Extra page 1. This page can be used for scratch work or as extra space. If you want work in the extra space (or on backs of pages) to be graded, indicate this very clearly on the page with the corresponding problem.

"expected number of BIRM jumpers"
among 3rd thru n^{th} jumpers

Let r.vs $I_3, I_4, \ldots, I_n$ indicate whether 3rd, 4th, ..., n^{th} jumper, respectively, was a BIRM jumper.

Y: counts BIRM jumpers among 3rd thru n^{th} jumpers!

$Y = \sum_{j=3}^{n} I_j$, so $E(Y) = E(\sum_{j=3}^{n} I_j) = \sum_{j=3}^{n} E(I_j)$.

Because $X_1, \ldots X_n$ are i.i.d w/ $\ln(X_i) \sim N(\mu, \sigma^2)$, each ordering (ranked) of jumpers is equally possible; SYMMETRY of i.i.d continuous r.v. So for <u>any</u> consecutive triplet of jumpers, it is equally likely that the third jumper's jump is the highest!

$E(I_j) = P(j^{th}$ jumper is BIRM$) = \frac{1}{3\cancel{!}}$. By SYMMETRY,

$E(Y) = (n-2)\frac{1}{3\cancel{!}}$

$\frac{1}{3}$, not $\frac{1}{3!} = \frac{1}{6}$, b/c

_ _ _ ← largest value in 3rd spot $\Rightarrow \frac{1}{3} = \frac{2!}{3!}$