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 ________________________________

Harvard ID ____________________________

1. ____________ (/12 points)

2. ____________ (/12 points)

3. ____________ (/12 points)

4. ____________ (/12 points)

Free ____2____ (/2 points)

Total ____________ (/50 points)

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)?

(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.)

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)?

(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.)

3. Let $X \sim \text{Bin}(100, 0.9)$. 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 $P(X \geq Y) = 1$?

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

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

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).

(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?

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

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.

Extra page 2. 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.

Table of Important Distributions

Let $0 < p < 1$ and $q = 1 - p$.

Name	Param.	PMF or PDF	Mean	Variance
Bernoulli	p	$P(X = 1) = p, P(X = 0) = q$	p	pq
Binomial	n, p	$\binom{n}{k} p^k q^{n-k}$, for $k \in \{0, 1, \dots, n\}$	np	npq
Geometric	p	$q^k p$, for $k \in \{0, 1, 2, \dots\}$	q/p	q/p^2
NegBinom	r, p	$\binom{r+n-1}{r-1} p^r q^n, n \in \{0, 1, 2, \dots\}$	rq/p	rq/p^2
Hypergeom	w, b, n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$, for $k \in \{0, 1, \dots, n\}$	$\mu = \frac{nw}{w+b}$	$(\frac{w+b-n}{w+b-1}) n \frac{\mu}{n} (1 - \frac{\mu}{n})$
Poisson	λ	$\frac{e^{-\lambda}\lambda^k}{k!}$, for $k \in \{0, 1, 2, \dots\}$	λ	λ
Uniform	$a < b$	$\frac{1}{b-a}$, for $x \in (a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	μ, σ^2	$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2

Stat 110 Midterm Solutions, Fall 2011

Prof. Joe Blitzstein

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 done with 20 randomly selected patients, the new treatment is effective for 15 of the patients.

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

Let B be the event that the new treatment is better than the standard treatment and let X be the number of people in the study for whom the new treatment is effective. By Bayes' Rule and the Law of Total Probability,

$$\begin{aligned}
P(B|X=15) &= \frac{P(X=15|B)P(B)}{P(X=15|B)P(B)+P(X=15|B^c)P(B^c)} \\
&= \frac{\binom{20}{15}(0.6)^{15}(0.4)^5(\frac{2}{3})}{\binom{20}{15}(0.6)^{15}(0.4)^5(\frac{2}{3})+\binom{20}{15}(0.5)^{20}(\frac{1}{3})}.
\end{aligned}$$

(This problem is a form of the "coin with a random bias" problem.)

(b) A second study is then done, with a set of 20 new random patients. Given the results of the first study, what is the conditional PMF for how many of the new patients the new treatment is effective on? (Don't simplify; it may help to let p be the answer to (a), and give your answer in terms of p.)

Let Y be how many of the new patients the new treatment is effective for nd $p = P(B|X=15)$ be the answer from (a). Then for $k \in \{0, 1, \ldots, 20\}$,

$$\begin{aligned}
P(Y=k|X=15) &= P(Y=k|X=15,B)P(B|X=15) + P(Y=k|X=15,B^c)P(B^c|X=15) \\
&= P(Y=k|B)P(B|X=15) + P(Y=k|B^c)P(B^c|X=15) \\
&= \binom{20}{k}(0.6)^k(0.4)^{20-k}p + \binom{20}{k}(0.5)^{20}(1-p).
\end{aligned}$$

(This distribution is *not* Binomial. As in the coin with a random bias problem, the individual outcomes are conditionally independent but not independent. Given the true probability of effectiveness of the new treatment, the pilot study is irrelevant and the distribution is Binomial, but without knowing that, we have a mixture of two Binomials.)

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)?

Similarly to the Strategic Practice problem about partnerships, imagine lining the people up and saying the first 3 are a team, the next 3 are a team, etc. This overcounts by a factor of $(3!)^{120} \cdot 120!$ since the order within teams and the order of teams don't matter. So the number of ways is

$$\frac{360!}{6^{120} \cdot 120!}.$$

(b) The 360 people consist of 180 married couples, and a random split into teams 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 the indicator for the jth team having a married couple (taking the teams to be chosen one at a time, or with respect to a random ordering). By symmetry and linearity, the desired quantity is $120E(I_1)$. We have

$$E(I_1) = P(\text{first team has a married couple}) = \frac{180 \cdot 358}{\binom{360}{3}},$$

since the first team is equally likely to be any 3 of the people, and to have a married couple on the team we need to choose a couple and then any third person. So the expected value is

$$\frac{120 \cdot 180 \cdot 358}{\binom{360}{3}}.$$

(This simplifies to $\frac{120 \cdot 180 \cdot 358}{360 \cdot 359 \cdot 358/6} = \frac{360}{359}$. Another way to find the probability that the first team has a married couple is to note that any particular pair in the team has probability $\frac{1}{359}$ of being married to each other, so since there are 3 disjoint possibilities the probability is $\frac{3}{359}$.)

3. Let $X \sim \text{Bin}(100, 0.9)$. For each of the following parts, construct an example showing that it is possible, or explain clearly why it is impossible.

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

This is impossible since there is a nonzero chance that Y is greater than 100, whereas X must be less than or equal to 100.

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

This is possible since we can have a sequence of trials where we "set the bar" for success at 2 different places. That is, X and Y are based on the same trials but with X having an easier threshold for success. For example, consider $U_1, \dots, U_{100}$ i.i.d. Unif(0,1), and define "success" on the jth trial to be $U_j \leq 0.9$ for X and $U_j \leq 0.5$ for Y. As another example, consider the chicken and egg homework problem, letting the probability of hatching be 0.9 and of hatching and surviving be 0.5, and let X be the number of chicks that hatch and Y be the number of chicks that survive.

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

This is impossible since if $Y - X \geq 0$ has probability 1, then $E(Y - X) \geq 0$ (an average of nonnegative numbers can't be negative!). But this would imply $E(Y) \geq E(X)$, contradicting $E(X) = 90, E(Y) = 50$. Intuitively, it would be ridiculous if a success probability of 0.5 rather than 0.9 on trials could somehow *guarantee* doing better or equal. Alternatively, note that $P(X \leq Y) = 1$ would imply $P(X = 0|Y = 0) = 1$, but we have

$$P(X = 0|Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} \leq \frac{P(X = 0)}{P(Y = 0)} < 1.$$

4. Athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with $X_1, X_2, \dots$ 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).

Let I_j be the indicator of the jth jumper being best in recent memory, for each $j \geq 3$. By symmetry, $E(I_j) = 1/3$ (similarly to examples from class and the homework). By linearity, the desired expected value is $(n-2)/3$.

(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?

The event $A_3 \cap A_4$ occurs if and only if the ranks of the first 4 jumps are $4,3,2,1$ or $3,4,2,1$ (where 1 denotes the best of the first 4 jumps, etc.). Since all orderings are equally likely,

$$P(A_3 \cap A_4) = \frac{2}{4!} = \frac{1}{12}.$$

(Alternatively, note that $A_3 \cap A_4$ occurs if and only if the 3rd and 4th jumpers set records, and apply the result of that homework problem.) As in (a), we have $P(A_3) = P(A_4) = 1/3$. So $P(A_3 \cap A_4) \neq P(A_3)P(A_4)$, which shows that A_3 and A_4 are not independent.

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

Let $Y_j = \ln(X_j)$, so that $X_j = e^{Y_j}$. By LOTUS,

$$\begin{aligned}\mathrm{Var}(X_j) &= E(e^{2Y_j}) - (E(e^{Y_j}))^2 \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{2y} e^{-(y-\mu)^2/(2\sigma^2)} dy - \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{y} e^{-(y-\mu)^2/(2\sigma^2)} dy \right)^2.\end{aligned}$$