Final exam review

You may use **three** sheets of notes (front and back) and a calculator on the final exam; otherwise same rules as for midterm exams. The final exam will be **Fri Dec 21, 9am-noon, in SMudd 014**.

Topics covered on the final exam:

- Properties of probabilities
- Finding probabilities for equally likely outcomes using counting principles
- Discrete distributions, including discrete uniform, Bernoulli, binomial, Poisson, geometric, negative binomial, hypergeometric, and multinomial
- Continuous distributions, including uniform, exponential, normal, gamma, beta, Pareto, and bivariate normal (and understand meaning of pdf and cdf)
- Poisson process
- Expectation and variance for discrete and continuous RVs
- Independence for discrete and continuous RVs
- Covariance and correlation for discrete and continuous RVs
- Joint distributions for discrete and continuous RVs
- Transformation methods (see sections 6.6, 6.10, 6.11, and 10.2)
- Conditional distributions for RVs
- Conditional expectation and variance for discrete and continuous RVs
- Bayes formula
- Laws of total expectation and total variance
- Law of Large Numbers
- Central Limit Theorem
- Moment-generating functions

Below are some more practice problems to help you prepare. You can also work through past STAT 360 exams (https://www.amherst.edu/academiclife/departments/mathematics-statistics/resources-opportunities/mathfinals/stat_360), though be aware that topics may vary. Good ones to work through are the Fall 2016 and Fall 2014 finals (skip the parts asking about R or Markov chains). In addition, work through all the problems handed out in class (also available on the course website) and past homework problems, and through the examples in the book.

Please let me know if you catch any errors in the brief solutions provided below.

1. The number of fish in a certain lake is a Pois(λ) random variable. Worried that there might be no fish at all, a statistician adds one fish to the lake. Let Y be the resulting number of fish (so Y is 1 plus a Pois(λ) random variable).
    a. Find $E[Y^2]$ and simplify.
    b. Find $E[1/Y]$ in terms of λ and simplify.

Answer: (a) $1+3\lambda+\lambda^2$ (b) $(1-e^{-\lambda})/\lambda$

2. Write the most appropriate of ≥ , ≤, =, or ? in the blank for each part (where "?" means that no relation holds in general). In (c) through (f), X and Y are i.i.d. (independent identically distributed) positive random variables. Assume that the various expected values exist.
    a. P(sum of 2 fair dice = 9) _____ P(sum of 2 fair dice = 10)
    b. P(65% of 20 children are girls) _____ P(65% of 2000 children are girls)
    c. $E[\sqrt{X}]$ _____ $\sqrt{E[X]}$
    d. E[sin(X)] _____ sin(E[X])
    e. P(X+Y > 4) _____ P(X>2)P(Y >2)
    f. $E[(X+Y)^2]$ _____ $2E[X^2] + 2(E[X])^2$

Answer: (a) ≥ (b) ≥ (c) ≤ (d) ? (e) ≥ (f) =

3. A fair die is rolled twice, with outcomes X for the 1st roll and Y for the 2nd roll.
    a. Compute the covariance of X+Y and XY.
    b. Are X+Y and XY independent? Justify your answer.
    c. Find the moment generating function $m_{X+Y}(t)$ of X+Y (your answer should be a function of t and can contain unsimplified finite sums).

Answer: (a) 245/12 (b) No, for example, X+Y=12 implies X=Y=6 so X-Y=0 (c) $\left(\frac{1}{6}\sum_{k=1}^{6} e^{kt}\right)^2$

4. A post office has 2 clerks. Alice enters the post office while 2 other customers, Bob and Claire, are being served by the 2 clerks. Alice is next in line. Assume that the time a clerk spends serving a customer has the Exp(λ) distribution.
    a. What is the probability that Alice is the last of the 3 customers to be done being served? Hint: no integrals are needed.
    b. Let X and Y be independent Exp(λ) RVs. Find the cdf of min(X,Y).
    c. What is the expected total time that Alice will spend at the post office (waiting plus being served)?

Answer: (a) ½ by memoryless property (b) min(X,Y) ~Exp(2λ) with cdf $F(z)=1-e^{-2\lambda z}$ (c) $3/(2\lambda)$

5. Let X and Y be independent standard normal RVs and let $R = \sqrt{X^2 + Y^2}$.
    a. Find the pdf of R.
    b. Find P(X > 2Y + 3) in terms of the standard normal cdf.
    c. Compute Cov(R^2, X). Are R^2 and X independent?

Answer: (a) $f(r)=re^{-r^2/2}$ for $r>0$ (b) $1-F(3/\sqrt{5}) = F(-3/\sqrt{5})$ (c) Cov = 0 but not independent (e.g., X large implies R large).

6. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $E(X_1) = 3$, and consider the sum $S_n = X_1 + X_2 + \cdots + X_n$.
   a. What is $E(X_1X_2X_3|X_1)$? (Your answer should be a function of X_1.)
   b. What is $E(X_1 \mid S_n) + E(X_2 \mid S_n) + \cdots + E(X_n \mid S_n)$?
   c. What is $E(X_1 \mid S_n)$? Hint: use (b) and symmetry.

Answer: (a) $9X_1$ (b) S_n (c) S_n/n

7. Let X and Y be independent Pois(λ) random variables. Recall that the moment generating function (mgf) of X is $m(t) = e^{\lambda(e^t-1)}$
   a. Find the mgf of X + 2Y.
   b. Is X + 2Y also Poisson? Show that it is, or that it isn't (whichever is true).

Answer: (a) $e^{\lambda(e^t+e^{2t}-2)}$ (b) No, either note mgf is not of Poisson form or find the expected value and variance to show they aren't equal, as they would be if Poisson.

8. An urn contains red, green, and blue balls. Balls are chosen randomly with replacement (each time, the color is noted and then the ball is put back.) Let r, g, b be the probabilities of drawing a red, green, blue ball respectively (r + g + b = 1).
   a. Find the expected number of balls chosen before obtaining the first red ball, not including the red ball itself.
   b. Find the expected number of different colors of balls obtained before getting the first red ball.
   c. Find the probability that at least 2 of n balls drawn are red, given that at least 1 is red.

Answer: (a) Geom(r) so expected number is (1-r)/r (b) g/(g+r) + b/(b+r) (c) $\frac{1-(1-r)^n-nr(1-r)^{n-1}}{1-(1-r)^n}$

9. Empirically, 49% of children born in the U.S. are girls (and 51% are boys). Let N be the number of children who will be born in the U.S. in January 2019, and assume that N is a Pois(λ) random variable, where λ is known. Assume that births are independent (don't worry about twins and such). Let X be the number of girls who will be born in the U.S. in January 2019, and let Y be the number of boys who will be born that month.
   a. Find the joint density of X and Y.
   b. Find E[N | X] and E[N^2 | X].

Answer: (a) $P(X = i, Y = j) = (e^{-0.49\lambda}(0.49\lambda)^i/i!)(e^{-0.51\lambda}(0.51\lambda)^j/j!)$ (b) $X+.51\lambda$, $(X+.51\lambda)^2+.51\lambda$

10. Let X_1, X_2, X_3 be independent with $X_i \sim \text{Exp}(\lambda_i)$ (with possibly different rates).
    a. Find $E[X_1+X_2+X_3 \mid X_1>1, X_2>2, X_3>3]$ in terms of λ_1, λ_2, λ_3.
    b. Find $P(X_1 = \min(X_1, X_2, X_3))$, the probability that the first of the three Exponential RVs is the smallest. Hint: re-state this in terms of X_1 and $\min(X_2,X_3)$; show $\min(X_2,X_3) \sim \text{Exp}(\lambda_2+\lambda_3)$.
    c. For the case $\lambda_1 = \lambda_2 = \lambda_3 = 1$, find the pdf of $\max(X_1, X_2, X_3)$.

Answer: (a) $1/\lambda_1 + 1/\lambda_2 + 1/\lambda_3+6$ (b) $\lambda_1/(\lambda_1+\lambda_2+\lambda_3)$ (c) $3(1-e^{-x})^2e^{-x}$

11. Joe's iPod has 500 different songs, consisting of 50 albums of 10 songs each. He listens to 11 random songs on his iPod, with all songs equally likely and chosen independently (so repetitions may occur).
    a. What is the pmf of how many of the 11 songs are from his favorite album?
    b. What is the probability that he listens to songs from more than one album (the 11 songs aren't all from the same album)? (Do not simplify.)

Answer: (a) $P(X=k) = \binom{11}{k}\frac{1}{50^k}\left(\frac{49}{50}\right)^{11-k}$ for $0\leq k \leq 11$ (b) $1-49!/(39!50^{10})$

12. A woman is pregnant with twin boys. Twins may be either identical or fraternal (non-identical). In general, 1/3 of twins born are identical. Obviously, identical twins must be of the same sex; fraternal twins may or may not be. Assume that identical twins are equally likely to be both boys or both girls, while for fraternal twins all possibilities are equally likely. Given the above information, what is the probability that the woman's twins are identical?

Answer: ½ (use Bayes formula with information that the twins are both boys)

13. A certain genetic characteristic is of interest. For a random person, this has a numerical value given by a $N(0, \sigma^2)$ RV. Let X_1 and X_2 be the values of the genetic characteristic for the twin boys from the previous problem. If they are identical, then $X_1 = X_2$; if they are fraternal, then X_1 and X_2 have correlation ρ. Find $\text{Cov}(X_1, X_2)$ in terms of ρ and σ^2.

Answer: $\sigma^2(1+\rho)/2$

14. Let X_1 be the number of emails received by a certain person today and let X_2 be the number of emails received by that person tomorrow, with X_1 and X_2 i.i.d.

   a. Find $E[X_1 \mid X_1 + X_2]$.
   b. For the case $X_j \sim \text{Pois}(\lambda)$, find the conditional distribution of X_1 given $X_1 + X_2$, i.e., $P(X_1 = k \mid X_1 + X_2 = n)$. What distribution is this?

Answer: (a) $(X_1+X_2)/2$ (b) Note $X_1 + X_2 \sim \text{Pois}(2\lambda)$ and use Bayes rule to get $\binom{n}{k}\left(\frac{1}{2}\right)^n$, which is Binom($n$,1/2)

15. Suppose X and Y are jointly distributed continuous random variables whose joint pdf is uniform over the region bounded by (0,0), (0,1), (2,2), and (2,1).

   a. Find the joint pdf of X and Y.
   b. Are X and Y independent?
   c. Find $P(Y < X)$.

Answer: (a) $f(x,y)=1/2$ for $0<x<1$, $x/2<y<x/2+1$, and 0 otherwise (b) No (c) ½

16. Let X and Y be jointly continuous random variables with joint pdf $f(x,y) = 6y$, for $0<y<x<1$, and 0 otherwise.

   a. Find the marginal density of Y. Be sure to specify the domain.
   b. Find the conditional pdf of X given Y.
   c. Calculate the conditional expectation of X given Y=0.25.

Answer: (a) $f_Y(y)=6y(1-y)$ for $0<y<1$ (b) $f_{X|Y}(x|y)=1/(1-y)$ for $0<y<x<1$ (c) $E[X|Y=0.25]=5/8$

17. Let X and Y be jointly distributed discrete RVs with joint pmf given in the following table:

Y \ X	0.3	0.6
0	0.1715	0.032
1	0.2205	0.144
2	0.0945	0.216
3	0.0135	0.108

   a. Find the marginal distribution of X.
   b. Determine $E[6X^2 + 19X]$.
   c. Find the conditional distribution of Y given that X = 0.6.
   d. Find $E[Y \mid X = 0.6]$.

Answer: (a) P(X=.3)=0.5=P(X=.6) (b) $E[6X^2 + 19X]=9.9$ (c) P(Y=0)=0.064, P(Y=1)=0.288, P(Y=2)=0.432, P(Y=3)=0.216 (d) 1.8

18. A writer with a music column asks his readers which types of jazz they like among dixieland, gypsy, and fusion. Results are that:

(i) 30% did not like any of these three types;
(ii) no readers like both dixieland and gypsy jazz;
(iii) the probability that a reader likes only dixieland is the same probability of liking only fusion;
(iv) the probability that a reader likes both dixieland and fusion is the same as the probability that a reader likes both gypsy and fusion, which is the same as the probability that a reader likes only gypsy
(v) the probability that a randomly selected reader likes jazz fusion is 0.4.

(a) Draw and label an appropriate Venn diagram describing the setting and the probabilities.

(b) What is the probability a reader likes dixieland?

(c) A reader tells the columnist that he prefers at least one of dixieland and fusion. Knowing that information, what is the probability that the reader likes exactly two of the types of jazz?

(d) Are the event that a reader likes dixieland and the event that a reader likes gypsy jazz mutually exclusive? independent?

Answer: (a) P(only dixieland)=0.2, P(only gypsy)=0.1 (b) 0.3 (c) 2/9 (d) yes; no

19. Lily is handing out treats to her cat Boots. She has been randomly generating X, the number of treats, using a binomial distribution with n=3 and p=0.2. She always wants to give at least one treat to Boots, so if 0 comes up, she gives out 1 treat instead. Let Y be this new RV: Y=X if X>0 and Y=1 if X=0.
   a. Make a table with the pmf of Y.
   b. Find the moment-generating function of Y.
   c. Use the mgf of Y to find E[Y].

Answer: (a) P(Y=1)=.896, P(Y=2)=.096, P(Y=3)=.008 (b) $m(t)=E[e^{tY}])=.896e^{t}+.096\,e^{2t}+.008\,e^{3t}$
(c) m'(0)=1.112=E[Y]

20. Suppose X is an exponentially distributed RV with mean 1/4. A researcher is interested in studying the random variable $Y = e^X$.
   a. Find the mean and variance of Y.
   b. State the Central Limit Theorem.
   c. What is the approximate distribution of $S_n=Y_1+...+Y_n$, where the Y_i are independent RVs with same distribution as Y and the sample size n is large?
   d. For n=50, the researcher observes a sample mean of 1.486. Is this value reasonable?

Answer: (a) E[Y]=4/3, V[Y]=2/9 (b) Check against p377 (c) S_n~Norm(4n/3,2n/9) (d) Sample mean is approx Norm(4/3,2/(9n)). For n=50, SD=1/15 and 1.486 is about 2.3 SDs from the expected value, which seems a bit unusual according to the 68-95-99.7 rule of thumb for normally distributed RV (less than 5% prob).

21. A fair coin is tossed five times. What is the probability of getting a sequence of at least three heads?

Answer: 1/4

22. If n balls are distributed randomly into k urns, what is the probability that the last urn contains j balls?

Answer: $\binom{n}{j}(k-1)^{n-j}/k^n$

23. A wine taster claims that she can distinguish four vintages of a particular Cabernet. What is the probability that she can do this by merely guessing? (She is confronted with four unlabeled glasses.)

Answer: 1/24

24. Urn A has three red balls and two white balls, and urn B has two red balls and five white balls. A fair coin is tossed. If it lands heads up, a ball is drawn from urn A; otherwise, a ball is drawn from urn B.
    a. What is the probability that a red ball is drawn?
    b. If a red ball is drawn, what is the probability that the coin landed heads up?

Answer: (a) 31/70 (b) 21/31

25. If U is uniform on [1,3], find the density function of $Y=U^2$. Be sure to specify the domain.

Answer: $f_Y(y) = 1/4\sqrt{y}$ for $1 \le y \le 9$

26. If X is uniform on (0, 1), and, conditional on X, Y is uniform on (0, X), find the joint and marginal distributions of X and Y.

Answer: $f(x,y) = 1/x$ for $0<y<x<1$ $f_X(x) = 1$ for $0<x<1$ $f_Y(y) = -\ln(y)$ for $0<y<1$

27. Three identical fair coins are thrown simultaneously until all three show the same face. What is the probability that they are thrown more than three times?

Answer: 27/64 (use negative binomial)

28. Suppose T is an exponential random variable with $P(T < 1) = .05$. What is λ?

Answer: Use the formula for the cdf of Exp(λ) to get $\lambda = -\ln(.95) \approx 0.0513$.

29. If $X \sim \text{Norm}(0,\sigma^2)$, find the density of $Y = |X|$.

Answer: $f_Y(y) = \frac{2}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)}$ for $y>0$ (using symmetry)

30. If X ~ Norm(μ,σ^2) and Y=aX+b, show that Y ~ Norm($a\mu+b$, $a^2\sigma^2$).

Answer: Can either show by directly working with the pdf, or use the mgf.

31. Suppose X and Y are independent standard normal variables. Let Z=X and W=X+Y.
   a. Use the Jacobian-based transformation formula to find the joint distribution of Z and W.
   b. Find the mean and variance for Z and for W.
   c. Find the correlation between Z and W.

Answer: (a) $f(z,w) = \frac{1}{2\pi}\exp(-\frac{1}{2}(z^2 + (w-z)^2))$ where z and w can be any real numbers (b) means are both zero, variances are 1 and 2 (c) correlation is $1/\sqrt{2}$

32. Suppose X and Y are independent standard normal variables. Let Z=Y/X. Show that Z has the Cauchy density, $f_Z(z) = \frac{1}{\pi(z^2+1)}$ for real numbers z.

Hint: Use a u-sub and symmetry to lead to an integral we know how to integrate.

33. Find the moment-generating function for a Bernoulli RV X, given probability p of success, and use it to calculate E[X], E[X^2] and E[X^3].

Answer: $m(t)=1-p+pe^t$ E[X]=p=E[X^2]=E[X^3]

34. Use the result of the previous problem to quickly deduce the mgf for Y ~ Binom(n,p). Use this mgf to find E[X].

Answer: $m(t)=(1-p+pe^t)^n$ $m'(0)=np$