December 9, 2019

OUTLINE OF MAJOR STAT 410 TOPICS

I. Probability Axioms, Sample Spaces, Set Manipulations, Conditional Probability, and Bayes' Rule.

II. Combinatorial Probability (counting problems, using binomial and hypergeometric building-blocks)

III. Discrete Random Variables (probabilities, expectations and distributions of functions, and joint distributions)

IV. Continuous Random Variables (same topics, plus distribution functions and hazards)

VI. Continuous density change-of-variable, uni- and multi-variate

VII. Joint-density probability calculations; checking of independence of joint random variables; expectations of functions of r.v.'s; covariance and correlation

VIII. Conditional density and expectation; properties of conditional expectation (conditioning and unconditioning)

IX. Moment Generating Functions; Densities and mass functions of sums of independent random variables; convolution formula; means and variances of sums of independent variables

X. Approximations, Limit Theorems (Poisson & Normal approximation to Binomial, LLN, CLT)

Sample Exam for Stat 410

Instructions: Do 200 points' worth of the following problems. (*The problems given below amount to quite a bit more than the 200 points worth that will be asked on the 2-hour exam.*) Your answers should be numerical expressions or functions of parameters (not integrals or summations), but you need not reduce your answers arithmetically or look up values of the standard normal distribution function $\Phi(\cdot)$. You may use a calculator and one two-sided 8×11 sheet of formulas and notes.

(1). Suppose $X_1, X_2, \ldots, X_{98}$ are independent and identically distributed continuous random variables with density $f(x) = 2\,x\,I_{[0 \le x \le 1]}$. Find approximate values for:

(a) *(15 points)* $P(63 \le X_1 + \cdots + X_{98} \le 70)$

(b) *(15 points)* P(no more than 50 of the rv's X_i, $1 \le i \le 98$, have values $\ge 3/4$)

(2). Recall that the *Negative Binomial* (r, p) is the distribution of the number of *Bernoulli*(p) trials required to achieve the r'th occurrence of 'Heads' or 'success'.

(a) *(20 points)* Find the moment generating function $M(t; r, p)$ of *Negative Binomial* as a (simplified) function of its arguments. *Use the identity based on the mass function:* $\sum_{n=r}^{\infty} \binom{n-1}{r-1} (1-x)^n = \left(\frac{1-x}{x}\right)^r$

(b) *(10 points)* Give a probabilistic interpretation or proof of the relation $M(t; r_1 + r_2, p) = M(t; r_1, p) \cdot M(t; r_2, p)$.

(3). *(25 points)* Find the joint density of $U = 2X + Y$, $V = X - Y$, and the marginal density of U, where $X \sim$ *Unif*$[0, 2]$ and $Y \sim$ Expon(1) are independent.

(4). *(25 points)* Suppose that it is known that 2% of all items manufactured on a certain assembly line are defective, and that from a warehouse containing 10,000 items manufactured on that line, 100 are randomly sampled **with replacement**. Give an exact expression for the probability that at most one of the items sampled was defective, and give an accurate simpler approximation for this probability.

(5). *(25 points)* Five cards are dealt at random (without replacement) from an ordinary deck of 52. Find the probability distribution (mass function) of the number of the cards dealt which are either picture cards (J, Q, or K) or from the spade suit.

(6). *(25 points)* Find the probability that the random variable $Y = X^3 + 2$ is at least 66, and find $E(Y)$, where X is a continuous random variable with density $f_X(x) = \frac{1}{36}\, x\,(6-x)\, I_{[0 \le x \le 6]}$.

(7). Suppose that U_1, U_2 are independent discrete random variables each taking on values in $\{0, \ldots, 9\}$ equiprobably. If $X = \max(U_1, U_2)$ and $Y = U_1 + U_2$, then find

(a) *(15 points)* the joint probability mass function of X, Y, and

(b) *(15 points)* the conditional expectation of X given $Y = 10$.

(8). *(20 points)* If $X_1, \ldots, X_{1000}$ are independent random variables with density $f_X(x) = 3x^2\, I_{[0<x<1]}$, then give the approximate values of

(a) $1000^{-1} \sum_{j=1}^{1000} I_{[X_j < 1/3]}$ and (b) $1000^{-1} \sum_{j=1}^{1000} 1/X_j$

(9). *(20 points)* You have 5 coins in your pocket, 3 of which are fair (i.e., have probability $1/2$ of showing heads on each toss), 1 of which has heads-probability $3/4$, and 1 of which has heads-probability $1/3$. One of the five coins is chosen at random and tossed independently 10 times, showing 5 heads and 5 tails. What is the probability that the coin chosen and tossed was one of the 3 fair ones ?

(10). Let X, Y have bivariate normal density given for all x, y by

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\Big(- \frac{1}{2(1-\rho^2)} (x^2 + y^2 - 2\rho x y) \Big)$$

(a) *(15 pts.)* Show that X and $Y - \rho X$ are independent normal variables with expectations 0, and use your argument to find their variances.
Hint: change-of-variable formula !

(b) *(15 pts.)* Show that $E(Y \mid X) = \rho X$, and use this to find the covariance and correlation of X, Y.

Solutions to Sample Exam Problems

(1). $E(X) = \int_0^1 2x^2\,dx = 2/3$, and $Var(X) = \int_0^1 2x^3\,dx - 4/9 = 2/4 - 4/9 = 1/18$. So $(X_1 + \cdots + X_{98} - 98(2/3))/\sqrt{98/18} \approx \mathcal{N}(0,1)$ by the CLT, so (a) = $\Phi((70-65.33)/(7/3)) - \Phi((63-65.33)/(7/3))$, and (b) = Binomial(98,p) probability of $[0,50]$ where $p = \int_{3/4}^1 2x\,dx = 7/16$, so the DeMoivre-Laplace approximation is $\Phi((50.5 - 98(7/16))/\sqrt{98(7/16)(9/16)})$.

(2). The probability mass function at n was the probability of $r-1$ successes in the first $n-1$ trials, followed by success in the n'th trial, or $\binom{n-1}{r-1}\, p^r\,(1-p)^{n-r}$. The indicated identity expresses just that for all $p = x$ the sum over $n \geq r$ of the probability mass function is 1. Then the mgf is

$$\sum_{n=r}^{\infty} \binom{n-1}{r-1}\, p^r\,(1-p)^{n-r}\, e^{tr} \;=\; (p/(1-p))^r \,\cdot\, ((1-p)e^t/(1-(1-p)e^t))^r$$

For part (b), the interpretation is that the sum of two independent $NegBin(r_1,p)$ and $NegBin(r_2,p)$ random variables, which is the waiting time until first r_1 and then r_2 more sucesses appear, has necessarily the same distribution as the $NegBin(r_1+r_2,p)$ waiting time for the (r_1+r_2)'th success.

(3). Since $u = 2x+y$, $v = x-y$ says the same thing as that $x = (u+v)/3$, $y = (u-2v)/3$, and $\det\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} = -3$, the Jacobian change-of-variable formula says directly that

$$f_{U,V}(u,v) \;=\; \frac{1}{2}\,e^{-(u-2v)/3}\,I_{[0<u+v<6,\,2v<u]}\,\frac{1}{3}$$

(Note that the two conditions on u in the indicator immediately imply that $u > 0$, which is the only condition on U alone.) Then the marginal density of U is obtained by integration: for each $u > 0$,

$$f_U(u) \;=\; \frac{1}{6}\int_{-u}^{\min(6-u,\,u/2)} e^{-(u-2v)/3}\,dv \;=\; \frac{1}{6}\,e^{-u/3}\,\frac{3}{2}\,\left[e^{\min(4-2u/3,\,u/3)} \;-\; e^{-2u/3}\right]$$

(4). Sampling with replacement implies $Binomial(100, 0.02)$ distribution, so the exact and approximate (Poisson(2)) probabilities are respectively

$$\sum_{k=0}^{1}\binom{100}{k}\,.02^k\,.98^{100-k} \;\approx\; 3\,e^{-2}$$

(5). The number of cards which are either pictures or spades is : $4(3)+13-3 = 22$. So the distribution is $\textit{Hypergeom}(52, 22, 5)$, which has probability mass function for $k = 0, \ldots, 5$:

$$p(k) \;=\; \binom{22}{k}\binom{30}{5-k} \,/\, \binom{52}{5}$$

(6). $P(Y > 66) \;=\; P(X^3 + 2 > 66) \;=\; P(X > 4) \;=\; \frac{1}{36}\int_4^6 x(6-x)dx \;=\; 1 - (80/3)/36 \;=\; 7/27$, and

$$EY \;=\; 2 + E(X^3) \;=\; 2 \,+\, \frac{1}{36}\int_0^6 x^4\,(6-x)\,dx \;=\; 2 \,+\, \frac{1}{36}\,((6/5)6^5 \,-\, (1/6)6^6)$$

(7). $p_{X,Y}(x,y) \;=\; \frac{1}{100}$ if $y = 2x$, and $= \frac{2}{100}$ if $0 \le x \le y \le 2x$, for $x = 0, \ldots, 9$ and $y = 0, \ldots, 18$. Now the marginal probability $P(Y = 10) \;=\; (1/10)\sum_{x=5}^{9}(2 - I_{[x=5]})/10 \;=\; 9/100$ and the conditional probability mass function is:

$$p_{X|Y}(x|10) \;=\; (100/9)\,\Big[(1/100)I_{[x=5]} \,+\, (1/50)I_{[x\in\{6,7,8,9\}]}\Big]$$

which leads to conditional expectation

$$E(X|Y=10) \;=\; 5(1/9) + (6+7+8+9)(2/9) = 65/9$$

(8). Here we apply the Law of Large Numbers, once to the independent identically distributed (iid) r.v.'s $I_{[X_j<1/3]}$ and once to the iid r.v.'s $1/X_j$, to obtain the answer $P(X_1 \,<\, 1/3) \;=\; \int_0^{1/3} 3x^2\,dx \;=\; 1/27$ in part (a), and $\int_0^1 3x^2\,(1/x)\,dx \;=\; 3/2$ in part (b).

(9). Bayes' rule gives as answer

$$(3/5)\binom{10}{5}2^{-10} \,/\, \Big[\binom{10}{5}\,\Big\{(3/5)\,2^{-10} \,+\, (1/5)\,3^5\,4^{-10} \,+\, (1/5)\,2^5\,3^{-10}\Big\}\Big]$$

or

$$3\,\cdot\,2^{-10} \,/\, \Big[3\,\cdot\,2^{-10} \,+\, 3^5\,\cdot\,4^{-10} \,+\, 2^5\,\cdot\,3^{-10}\Big] \;=\; 0.791$$

(10). The change of variable is $(x,y) \mapsto (x,\,w)$ with $w \,=\, y - \rho x$, so that $y = w + \rho x$ and the Jacobian is the matrix with first row $(1,0)$ and second row $(-\rho, 1)$, and therefore has determinant 1. If $W \,=\, Y - \rho X$, then the change of variables formula gives

$$f_{X,W}(x,w) \;=\; \frac{1}{2\pi(1-\rho^2)}\,\exp\Big(-\,\frac{x^2}{2}\,-\,\frac{w^2}{2(1-\rho^2)}\Big)$$

Therefore, since the joint density factors, $X \sim \mathcal{N}(0,1)$ and $W \sim \mathcal{N}(0, 1-\rho^2)$ are independent, and $E(Y - \rho X \mid X) = E(W \mid X) = E(W) = 0$, which implies $E(Y \mid X) = \rho X$ and $E(XY) = E(XE(Y \mid X)) = \rho E(X^2) = \rho$. It follows that $\mathrm{Cov}(X,Y) = \rho$, and $\mathrm{Corr}(X,Y) = \rho/\sqrt{1-\rho^2}$.