MATH 525 | FINAL EXAM | FALL 2004

This is a 2-hour exam. You may use three letter-size sheets with notes.

Show all your work!

Good luck!

Name : ____

Facts that may be useful:

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \ , \quad |q| < 1 \ ; \qquad \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x \ ; \qquad (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$\lim_{n\to\infty} \left(1 + \frac{\alpha}{n}\right)^n = e^{\alpha}$$

Some important distributions:

- If $X \sim \text{Bin}(n,p)$, then $f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \ldots, n$; $\mathbb{E}X = np$, $\text{var}X = np(1-p)$; $\phi(t) = (1-p+pe^{it})^n$.
- If $X \sim \text{Geometric}(p)$, then $f_X(k) = (1-p)^{k-1}\, p$ for $k \in \mathbb{N}$; $\mathbb{E}X = \frac{1}{p}$, $\text{var}X = \frac{1-p}{p^2}$.
- If $X \sim \text{Poisson}(\lambda)$, then $f_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k \in \{0,1,2,\ldots\}$; $\mathbb{E}X = \lambda$, $\text{var}X = \lambda$; $\phi(t) = \exp[\lambda(e^{it}-1)]$.
- If $X \sim \text{Uniform}(a,b)$, then $f_X(x) = \frac{1}{b-a}$ for $x \in [a,b]$, and 0 otherwise; $\mathbb{E}X = \frac{a+b}{2}$, $\text{var}X = \frac{(b-a)^2}{12}$.
- If $X \sim \text{Exp}(\lambda)$, then $f_X(x) = \lambda e^{-\lambda x}$ for $x \in [0,\infty)$, and 0 otherwise; $\mathbb{E}X = \frac{1}{\lambda}$, $\text{var}X = \frac{1}{\lambda^2}$.
- If $X \sim \text{N}(\mu,\sigma^2)$, then $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$; $\mathbb{E}X = \mu$, $\text{var}X = \sigma^2$.

Problem 1.

This problem is about the hypergeometric random variables and its relationship with the binomial and Poisson random variables. I will formulate it in terms of elk in a forest.

Let there be N elk in a forest, and let M of them be "spotted" elk, i.e., a sub-species of elk that has spots on the skin. (I am not sure that one can talk about "spotted" elk, but that's probability, not biology.) Assume that the spotted and non-spotted type of elk have the same chance of being captured by us. Now assume that we capture n elk, and that each particular elk is captured independently of the other elk. Let X be the number of spotted elk among the n captured ones. Then X is a *hypergeometric* random variable, and its p.m.f. is

$$f_X(k) = \mathbb{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \,, \qquad k = 0,\, 1,\, \ldots,\, n \,.$$

Recall that to derive this formula, we think as follows: to have k spotted elk out of n captured elk, we have to choose k out of total of M spotted elk in the forest, and $(n-k)$ out of total of $(N-M)$ non-spotted ones.

(a) Assume that the total elk population of the forest, N, is very large, and that the proportion, $\frac{M}{N}$, of the spotted elk is small. Also, assume that the number n of captured elk is much smaller than M and N. Then the distribution of X can be *approximated* by binomial distribution. Explain intuitively this fact, and find the parameters of this binomial distribution. (I don't want you to give a formal proof, but just to explain in a couple of sentences.)

(b) In a realistic situation, we do not know the total elk population, N, as well of the number M of spotted elk, but only the proportion, $p = \frac{M}{N}$, of spotted elk. Assume that p is small, i.e., that spotted elk are quite rare. We capture n elk, where n is such that the product np is moderate. As before, X is the random variable equal to the number of spotted elk among the captured n elk. In part (a), we found that X can be approximated by a binomial random variable. Under conditions like the ones just described, a binomial random variable can be approximated by Poisson random variable. Use the characteristic functions of binomial and Poisson distributions to prove that in the limit of very large number n of captured elk, very small proportion spotted elk, and moderate value of np, the distribution of a binomial(n, p) random variable tends to a Poisson distribution.

Problem 2.

Negative binomial random variables occur in the following situation. Let us perform a sequence of independent Bernoulli trials, each one having probability of success p (and, therefore, probability of failure $1-p$). Let r be a positive integer. We say that the random variable X is *negative binomial* with parameters r and p, and write $X \sim \mathrm{NegBin}(r,p)$, if X is the number of trials needed to get r successes. Note that a random variable of type $\mathrm{NegBin}(1,p)$ is the same as what we called geometric random variable with parameter p.

The p.m.f. of $X \sim \mathrm{NegBin}(r,p)$ is

$$f_X(k) = \mathbb{P}(X=k) = \binom{k-1}{r-1} p^k (1-p)^{k-r} , \qquad k = r,\, r+1,\, r+2,\, \dots ,$$

and the moment generating function of X is

$$M_X(t) = E\left[e^{tX}\right] = \left(\frac{p}{1-(1-p)\,e^t}\right)^r .$$

Let $X \sim \mathrm{NegBin}(r,p)$ and $Y \sim \mathrm{NegBin}(s,p)$ be independent random variables. You have to prove that $X+Y \sim \mathrm{NegBin}(r+s,p)$. You can do this in at least three ways, and each correct proof (or explanation) will bring you additional points.

(a) *Method 1.* Just think about the meaning of the random variable $X+Y$.

(b) *Method 2.* Find the moment generating function $M_{X+Y}(t)$ of the random variable $X+Y$, and identify the type of $X+Y$. (That's the easiest method!)

(c) *Method 3.* Using the p.m.f. of negative binomial random variables (given above), you can prove that

$$p_{X+Y}(k) = \binom{k-1}{r+s-1} p^k \, (1-p)^{k-(r+s)} \ , \qquad k = r+s, \, r+s+1, \, r+s+2, \, \ldots \ .$$

To do this, assume that $r < s$ (which does not restrict the generality of your proof), and use the identity

$$\sum_{m=r}^{k} \binom{m-1}{r-1} \binom{k-m-1}{s-1} = \binom{k-1}{r+s-1} \ ,$$

which you can take for granted. If you insist on giving a combinatorial interpretation of this identity, you can try – of course, it will bring you a couple of points, but **PLEASE** do not attempt to do this before you finish all other problems!!!

Problem 3.

If $X \sim \mathrm{Bin}(n,p)$, and $Y \sim \mathrm{NegBin}(r,p)$, explain in a couple of sentences why

$$\mathbb{P}(X < r) = \mathbb{P}(Y > n) \ .$$

Problem 4.

Adam and Bill take turns rolling a fair die – Adam rolls the die first, then Bill, then Adam again, and so on. The first to roll a 6 wins the game. Find the probability that Adam wins the game.

If you can solve the problem in two different ways, do so – you will get twice as many points.

Method 1: You can calculate the probability that Adam wins as the sum of the disjoint events A_1, A_3, A_5, $\ldots$, where A_n is the event that Adam wins in the nth round. You will have to use the formula for geometric series in the form $1 + q^2 + q^4 + q^6 + \cdots = \frac{1}{1-q^2}$ valid for $|q| < 1$.

Method 2: You can condition on the outcome of the first two rolls, and use the law of total probability. That is, let i_1, i_2, i_3, etc., be the outcomes of the 1st, 2nd, 3rd, etc., rolls; i_1 can be 6 (if the 1st roll is 6) or $\bar{6}$ (if the 1st roll is not 6), and similarly for i_2, i_3, etc. With this notation, the outcome of the first two rolls can be written as (i_1, i_2).

Problem 5.

The number R of people who enter an elevator on the ground floor of a building is a Poisson random variable with mean 10. If there are N floors above the ground floor (the ground floor does not count!), and if each person is equally likely to get off the elevator at any one of these N floors, independently of where the other people get off, compute the expected number of stops that the elevator will make before discharging all of its passengers.

Note that the elevator does not stop at the floors where nobody wants to get off, and there are no people outside the elevator.)

Let A_i be the event "somebody gets off at floor $\#i$", and let I_{A_i} be its indicator function.

(a) Express the number of stops, S, in terms of the indicator functions of the events A_i.

(b) Find $\mathbb{E}[S]$ in terms of the expectation of A_1.

Hint: That's *very* easy.

(c) If $R = r$, what is the probability that *none* of the r people that were in the elevator initially (at the ground floor) will get off at floor #1?

(d) Show that the conditional expectation $\mathbb{E}[I_{A_1}|R = r]$ is equal to $1 - \left(\frac{N-1}{N}\right)^r$.

Hint: This is not difficult if you think about the meaning of all the random variables. Also, recall that the probability of an event is related to the expectation of the indicator function of the event.

(e) Use (d) and your favorite theorem ($\mathbb{E}Y = \mathbb{E}[\mathbb{E}(Y|X)]$) to find $\mathbb{E}[I_{A_1}]$.

(f) Finally, put together (b) and (e) to show that $\mathbb{E}[S] = N\left(1 - e^{-10/N}\right)$.

Problem 6.

Let U_1, U_2, $\ldots$ be a sequence of independent Uniform$(0, 1)$ random variables. Let $x \in (0, 1]$, and

$$N(x) := \min \left\{ n \in \mathbb{N} \ : \ \sum_{j=1}^{n} U_i > x \right\}$$

be the smallest number of U's that need to be added so that their sum exceeds x. From the definition, it is clear that $N(x)$ is a random variable taking values in $\mathbb{N}$.

(a) Explain why $\mathbb{P}\big(N(x) \geq 1 \big) = 1$.

(b) Explain why

$$\mathbb{P}\big(N(x) \geq n + 1 \mid U_1 = u_1 \big) = \mathbb{P}\big(N(x - u_1) \geq n \big) \qquad \text{for } x > u_1 \ .$$

Hint: Look at the picture above.

(c) Show by induction on n that

$$\mathbb{P}\big(\, N(x) \geq n+1 \,\big) = \frac{x^n}{n!} \;, \qquad n \in \{0, 1, 2, \dots\} \;.$$

Note that in part (a) you already proved that this relationship holds for $n = 0$.

Hint: Use conditioning on U_1, i.e.,

$$\mathbb{P}\big(\, N(x) \geq n+1 \,\big) = \mathbb{E}\left[\mathbb{P}\big(\, N(x) \geq n+1 \;\big|\; U_1 \,\big)\right] \;.$$

This equation works in the same way as your favorite theorem; naturally, the averaging here is over U_1. In the integral you obtain, apply (b) and the induction hypothesis, and be careful with the limits of integration.

(d) Use part (c) and the formula

$$\mathbb{E}[Y] = \sum_{n=0}^{\infty} \mathbb{P}\big(Y \geq n+1 \big) \tag{$*$}$$

which holds for any random variable Y taking values in $\{0, 1, 2, \ldots\}$, to find $\mathbb{E}[N(x)]$. If you also prove $(*)$, you will get two bonus points.

The End!