Last Name: *First Name:* *Student Number:*

UNIVERSITY OF TORONTO
Faculty of Arts and Science
DECEMBER EXAMINATIONS 2004
STA 247H1 F
Duration - 3 hours

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
T	

Answer all questions in the space provided; if you run out of space, use the back of a page.

No books or notes are allowed. No calculators are allowed.

You may use the facts about standard distributions and the table of normal probabilities that are on the sheets at the end of this exam.

For all questions requiring a numerical answer, you must produce an actual number as your answer — ie, a decimal number (eg, 0.15) or a simple fraction (eg, 3/17).

For all questions, show how you obtained your answer (except when the question says no explanation is necessary).

The ten questions are worth equal amounts.

1. You flip a nickel three times and a quarter three times. Define the following events:

A	The event that the nickel lands heads the first time it is flipped
B	The event that the quarter lands heads the first time it is flipped
C	The event that the nickel lands heads three times
D	The event that the nickel and the quarter land heads the same number of times
E	The event that the nickel lands heads fewer times than the quarter lands heads

(a) Compute $P(C)$.

(b) Compute $P(C|A)$.

(c) Compute $P(D)$.

(d) Compute $P(C|D)$.

(e) Are A and B independent? Answer yes or no (no explanation is necessary).

(f) Are A and C independent? Answer yes or no (no explanation is necessary).

(g) Are A and D independent? Answer yes or no (no explanation is necessary).

(h) Are A and E independent? Answer yes or no (no explanation is necessary).

2. Suppose that a technical support person takes 6 minutes on average to deal with one customer's problem, that the standard deviation of the time required to deal with one customer is 2 minutes, and that it never takes more than 15 minutes to deal with one customer. Suppose that each customer's problem is independent of the problems of other customers. Find a good approximation to the probability that the technical support person will take more than 175 minutes to deal with the problems of 25 customers.

3. Let X and Y be independent random variables. X has the geometric distribution with parameter $p = 1/2$. Y has the binomial distribution with $n = 8$ and $p = 1/2$. Prove that $P(X + Y > 12)$ is less than $1/8$.

4. We have studied the following families of distributions in this course:

Bernoulli binomial geometric negative binomial Poisson exponential normal

For each of the following situations, say which of these families of distributions might be appropriate for modeling the quantity of interest, and explain briefly why it might be appropriate (ie, say why a distribution in this family might be the correct one, or a good approximation to the correct one). Also, say what additional assumptions (if any) would be needed to justify using the distribution. If more than one distribution would be appropriate, name all those that would be. If none of the distributions above would be appropriate, say that (no explanation is needed in this case). Note: The information given may not be sufficient to determine the parameters of the distribution. We're interested here only in which family (or families) would be appropriate.

(a) For your nineteenth birthday, an uncle says that he will buy you tickets for Blue Jays baseball games. Specifically, he says that he will pay for tickets to baseball games until you have been to three games in which the total runs scored (by both teams) is nineteen. You are interested in modeling the distribution of the number of baseball games he will pay for.

(b) Every evening, a company runs a program on the computer system used to manage energy usage in their buildings that analyses how much energy was used that day, and produces a report of energy usage for managers to read. The computer system does nothing else while running this program. The operations the program performs are the same every day, but nevertheless the time it takes to run varies slightly from day to day, due to various factors, such as slight variations in the rotational speed of the hard disks (which affects transfer rates), different timing of device interrupts, which changes what data is in cache memory at different times, and the occasional need to re-send network packets when they weren't received correctly the first time. You are interested in modeling the time it takes for this program to run.

(c) A tax auditor examines income tax returns, looking for ones in which the taxpayer has incorrectly claimed more than their allowed deduction for RSP contributions (either because they made a mistake, or because they are trying to cheat). Only a few taxpayers do this (less than one in a hundred). The tax auditor always examines 1000 tax returns in a day. You are interested in modeling the distribution of the number of tax returns that the tax auditor examines in one day that have an incorrect RSP deduction.

(d) A train company's trains usually arrive nearly on time (sometimes slightly early, sometimes slightly late), but they occasionally are delayed by a variety of problems, such as trees having fallen on the tracks, or mechanical problems with the locomotive. You are interested in modeling the distribution of how late a train is (which will be negative if the train is early).

(e) A company is interested in how often people who access its home web page click on the link that explains a special sales offer. They look at the next 100,000 accesses to their home web page, and record the number of these access that were followed by a click on this link by the same person. You are interested in modeling the distribution of the number of these 100,000 accesses in which the person clicked on the link.

5. Suppose that the random variable X has the following probability mass function:

x	0	1	2	3	4
$p(x)$	0.03	0.64	0.32	0	0.01

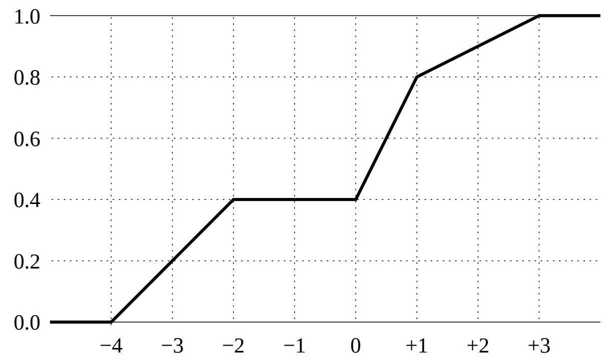
Suppose also that the conditional distribution of Y given $X = x$ is binomial with parameters $n = 3$ and $p = X/4$.

(a) Write down a table showing the joint probability mass function for X and Y.

(b) Find the conditional probability mass function of X given $Y = 3$.

(c) Find the expected value of Y.

6. Here is a graph of the cumulative distribution function for a random variable X:



(a) Draw a graph of the probability density function for this random variable. Label the important points on the horizontal and vertical axes to make clear exactly what the function is like.

(b) Find $P(-3 \leq X \leq 1)$.

(c) Compute $E(X)$.

7. You roll 36 fair, six-sided dice. You then remove the dice that show six. Let X be the number of dice remaining at this time. You roll these remaining dice a second time. Again, you remove the ones that show six. Let Y be the number of dice remaining at this time. Find the numerical values of each of the following, and briefly explain how you obtained your answer.

   (a) $E(X)$.

   (b) $\mathrm{VAR}(X)$.

   (c) $E(Y|X = 24)$.

   (d) $E(Y)$.

   (e) $\mathrm{VAR}(Y)$.

8. Let X and Y be independent random variables. X has the exponential distribution with parameter λ_X (and hence has mean $1/\lambda_X$). Y has the exponential distribution with parameter λ_Y. Define the random variable Z to be the minimum of X and Y. Prove that Z has the exponential distribution with parameter $\lambda_X + \lambda_Y$.

9. A government bureaucrat has the job of processing applications for hunting licences. At the start of the day, some number of licence applications arrive at his desk, and are added to the pile of licence applications left unprocessed from previous days. The bureaucrat can process 30 applications in a day. However, if the number of applications waiting for him to process at the start of the day (after the new ones arrive) is less than 50, the bureaucrat takes a long lunch break, and consequently processes only 15 applications that day.

   Suppose that the number of applications arriving each morning has the Poisson distribution with mean 20, and that the number of applications arriving one day is independent of the number arriving any other day. Write an R function that simulates the bureaucrat doing his job for 1000 days, and that estimates the probability that the bureaucrat will take a long lunch break once the situation has reached a steady state. Assume that on the first day, no applications are waiting to be processed.

   To generate a random value from the Poisson distribution with mean `lambda`, use `rpois(1,lambda)`.

10. Suppose that where you live, the weather doesn't depend on the season. Suppose also that where you live, the weather tomorrow depends only on what the weather is today — ie, if you know the weather today, knowing the weather on earlier days wouldn't help you predict the weather tomorrow. Suppose also that if it rains today, the probability of it raining tomorrow is $1/2$, and if it doesn't rain today, the probability of it raining tomorrow is $1/4$. In the long run, how often does it rain where you live (ie, on what fraction of days does it rain)?

Facts about standard distributions

Binomial distribution

Parameters are n and p. Range is the integers from 0 to n.

Probability mass function: $p(x) = \binom{n}{x} p^x(1-p)^{n-x}$

Mean: $E(X) = np$
Variance: $\text{Var}(X) = np(1-p)$

Geometric distribution

Parameter is p. Range is the integers from 1 on up.

Probability mass function: $p(x) = p(1-p)^{x-1}$

Mean: $E(X) = 1/p$
Variance: $\text{Var}(X) = (1-p)/p^2$

Negative binomial distribution

Parameters are k and p. Range is the integers from k on up.

Probability mass function: $p(x) = \binom{x-1}{k-1} p^k(1-p)^{x-k}$

Mean: $E(X) = k/p$
Variance: $\text{Var}(X) = k(1-p)/p^2$

Poisson distribution

Parameter is μ. Range is the integers from 0 on up.

Probability mass function: $p(x) = e^{-\mu}\mu^x/x!$

Mean: $E(X) = \mu$
Variance: $\text{Var}(X) = \mu$

Exponential distribution

Parameter is λ. Range is the positive real numbers.

Probability density function: $f(x) = \lambda e^{-\lambda x}$

Mean: $E(X) = 1/\lambda$
Variance: $\text{Var}(X) = 1/\lambda^2$

Normal distribution

Parameters are μ and σ. Range is the real numbers.

Probability density function: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$

Mean: $E(X) = \mu$
Variance: $\text{Var}(X) = \sigma^2$