Summary of the standard discrete densities

Bernoulli(p): X=result of a single Bernoulli trial.

$$f(k) = P(X = k) = \begin{cases} 1-p & \text{if } k = 0 \\ p & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases},$$

$E[X] = p, \quad \mathrm{Var}(X) = p(1-p).$

Binomial(n, p): X=number of successes in n independent Bernoulli trials.

$$f(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases},$$

$E[X] = np, \quad \mathrm{Var}(X) = np(1-p).$

Note. Binomial($1, p$)=Bernoulli(p).

Geometric(p): X=number of trials for the first success, in an infinite sequence of Bernoulli trials.

$$f(k) = P(X = k) = \begin{cases} p(1-p)^{k-1} & \text{if } k = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases},$$

$E[X] = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$

Useful formula: $P(X \geq k) = (1-p)^{k-1}$

Negative Binomial(r, p): X=number of trials for achieving r successes.

$$f(k) = P(X = k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r} & \text{if } k = r, r+1, r+2, \ldots \\ 0 & \text{otherwise} \end{cases},$$

$E[X] = \frac{r}{p}, \quad \mathrm{Var}(X) = r\frac{1-p}{p^2}.$

Note. Negative Binomial($1, p$)=Geometric(p).

Poisson(λ): X=number of events occurring during a time interval of length 1, given that events occur at an average rate of λ per unit time.

$$f(k) = P(X = k) = \begin{cases} e^{-\lambda}\lambda^k/k! & \text{if } k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases},$$

$E[X] = \lambda, \quad \mathrm{Var}(X) = \lambda.$

Note that $X \sim$ Poisson(λT) is used for the number of events occurring during a time interval of length T, given that events occur at an average rate of λ per unit time.

Remark. $X \sim$ Poisson(λ) with $\lambda = np$ provides a good approximation to Binomial(n, p), assuming n is large and np is moderate.

Hypergeometric(n, N, m): X=number of white balls in a sample of n balls taken without replacement from an urn of N balls, of which m are white and $N - m$ are black.

$$f(k) = P(X = k) = \begin{cases} \binom{m}{k}\binom{N-m}{n-k} \big/ \binom{N}{n} & \text{if } k = 0, 1, 2, \ldots, n, \\ 0 & \text{otherwise,} \end{cases}$$

$E[X] = n\frac{m}{N}, \quad \mathrm{Var}(X) = \frac{N-n}{N-1}\, n\frac{m}{N}\left(1 - \frac{m}{N}\right).$

Summary of the standard continuous densities

Uniform(a,b): X=randomly chosen point in the interval (a,b). Density

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise,} \end{cases}$$

$E[X] = \frac{b+a}{2}$, $\quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$.

Normal(μ,σ^2): X=random fluctuation arising from many causes. Density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty,$$

$E[X] = \mu$, $\quad \mathrm{Var}(X) = \sigma^2$. When $\mu = 0$ and $\sigma = 1$, one obtains the standard normal density $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Useful formula: $P(a \le X \le b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$ where Φ is the distribution function of the standard normal density: $\Phi(z) = \int_{-\infty}^{z} \phi(x)\,dx$.

Important result: the Central Limit Theorem (Section 8.3) says that the distribution of outcomes is approximately normal, after many independent repetitions of a random experiment, no matter what the experiment is!

Exponential(λ): X=waiting time until the first event, if there are λ events per unit time on average. Density

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0, \\ 0 & \text{otherwise,} \end{cases}$$

$E[X] = \frac{1}{\lambda}$, $\quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$.

Useful formula: $P(X > x) = e^{-\lambda x}$ for $x \ge 0$.

(*Discrete analogue* of Exponential(λ) is Geometric(p) with $p = 1 - e^{-\lambda}$.)

Gamma(α,λ): X=waiting time until the α^{th} event (when α is a positive integer), if there are λ events per unit time on average. Density

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & \text{if } x \ge 0, \\ 0 & \text{otherwise,} \end{cases}$$

$E[X] = \frac{\alpha}{\lambda}$, $\quad \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$.

Note. Gamma$(1,\lambda)$=Exponential(λ).

Cauchy: X=tangent of a uniformly randomly chosen angle between $-\pi/2$ and $\pi/2$. Density

$$f(x) = \frac{1}{\pi(1+x^2)},$$

$E[X] = 0$, $\quad \mathrm{Var}(X) = +\infty$.

Exercise. Sketch each of the density functions $f(x)$, to get a feel for which X-values are more likely to occur, in each case.

NAME:

Math 461 Spring 2005 — Test 1

Total points: **100**. Do all questions. **Explain** all answers. No notes, books, or electronic devices.

1. [15 points] Fifteen distinct balls are randomly divided among players E, F and G, with each player getting five balls.

If ten of the balls are White and five are Black, then find the probability player E gets 2 White balls, player F gets 3 White balls, and player G gets 5 White balls.

2. [30=10+10+10 points] Suppose n identical balls are to be distributed into 4 boxes, with at least one ball in each box.

(a) Explain why this can be done in $\binom{n-1}{3}$ ways.

(b) Assume each distribution is equally likely. Find the probability that the first box has just one ball in it.

(c) Assume $n \geq 5$. Find the probability that at least one box has just one ball in it.

3. [15=5+5+5 points]
(a) Randomly choose k cards without replacement from a deck of 52 cards. Find the probability that the cards contain no aces.

(b) Instead suppose you choose k cards with replacement (meaning you randomly choose a card, make a note of what it is, then replace it, and repeat). Find the probability that the cards contain no aces.

(c) For a certain unfair die, we know the probability of rolling the number k equals $\frac{k}{21}$ (for $k = 1, 2, 3, 4, 5, 6$).

First roll the die and then choose k cards with replacement, from the deck. Find the probability that the chosen cards contain no aces.

4. [10 points] A parallel system functions whenever at least one of its components works. Consider a parallel system with $n \geq 2$ components and suppose that each component independently works with probability $\frac{2}{3}$.

Assuming the system is functioning, find the probability that the first two components are working.

5. [20 points] People come in 10 types, with a randomly chosen person having probability $\frac{j}{55}$ of being type $j = 1, 2, \ldots, 10$. The probability that a person of type j likes wine is $\frac{j-1}{45}$.

Given that a person likes wine, what is the probability that the person is of type 3?

6. [10 points] A hiker is missing and is thought to be in one of two regions: Region 1 with probability $\frac{1}{3}$ and Region 2 with probability $\frac{2}{3}$. The probability that a search of Region 1 finds the hiker is $1 - \beta_1$ if the hiker is in Region 1, and the probability that a search of Region 2 finds the hiker is $1 - \beta_2$ if the hiker is in Region 2.

A search of Region 1 has failed to find the hiker.

(a) Find the probabilities that the hiker is in Region 1 and Region 2, respectively.

(b) Which region should be searched next, to have the best chance of finding the hiker? (The answer depends on β_1 and β_2.)

Math 461 Spring 2005 — Test 1 SOLUTIONS

Total points: **100**. Do all questions. **Explain** all answers. No notes, books, or electronic devices.

1. [15 points] Fifteen distinct balls are randomly divided among players E, F and G, with each player getting five balls.

If ten of the balls are White and five are Black, then find the probability player E gets 2 White balls, player F gets 3 White balls, and player G gets 5 White balls.

Solution 1, without ordering. If the White balls are distributed as above, then (by subtraction) player E gets 3 Black balls, player F gets 2 Black balls, and player G gets 0 Black balls. So put the White balls into groups of size 2,3,5 and the Black balls into groups of size 3,2,0, then multiply these possibilities together and divide by the total number of ways of putting 15 balls into groups of size 5,5,5:

$$\text{probability} = \frac{\binom{10}{2,3,5}\binom{5}{3,2,0}}{\binom{15}{5,5,5}}.$$

Solution 2, with ordering. Imagine 15 positions numbered 1–15, with positions 1–5 reserved for player E, positions 6–10 for player F, and positions 11–15 for player G. Let us count how many ways the balls can be arranged in the desired fashion:

$$\binom{5}{2}\binom{5}{3}\binom{5}{5}10!\,5!,$$

where we first choose the positions for the White balls, which of course determines the positions of the Black balls, and then permute the White balls and Black balls amongst themselves.

For the denominator, the balls can be put in position in 15! ways. So the answer is

$$\text{probability} = \frac{\binom{5}{2}\binom{5}{3}\binom{5}{5}10!5!}{15!}.$$

You can check that this equals the answer in Solution 1.

2. [30=10+10+10 points] Suppose n identical balls are to be distributed into 4 boxes, with at least one ball in each box.

(a) Explain why this can be done in $\binom{n-1}{3}$ ways.

Solution. Lay the n balls out in a line, and then insert 3 "dividers" into some of the $n-1$ spaces between balls. This can be done in $\binom{n-1}{3}$ ways, and this procedure divides the balls into 4 groups, to be put in the 4 boxes.

(b) Assume each distribution is equally likely. Find the probability that the first box has just one ball in it.

Solution. If the first box has just one ball in it, then $n-1$ balls remain to be distributed among the other 3 boxes. This can be done in $\binom{n-2}{2}$ ways, by arguing like in part (a) (inserting 2 dividers in some of the $n-2$ spaces between balls). Hence the probability of this happening is

$$\frac{\binom{n-2}{2}}{\binom{n-1}{3}} = \frac{3}{n-1}.$$

(c) Assume $n \geq 5$. Find the probability that at least one box has just one ball in it.

Solution 1.

$$\begin{aligned} P(\text{at least one box has just one ball in it}) &= 1 - P(\text{no box has just one ball in it}) \\ &= 1 - P(\text{every box has} \geq 2 \text{ balls in it}) \\ &= 1 - \frac{\binom{n-5}{3}}{\binom{n-1}{3}}, \end{aligned}$$

since the way to ensure each box has ≥ 2 balls is to first put one ball in each box and then distribute the remaining $n-4$ balls among the 4 boxes (similar to part (a)).

Solution 2. Let A_k =event that box k has just one ball. Then

$$
\begin{aligned}
&P(\text{at least one box has just one ball in it})\\
&= P(\cup_{k=1}^4 A_k)\\
&= \sum_k P(A_k) - \sum_{k_1<k_2} P(A_{k_1}A_{k_2}) + \sum_{k_1<k_2<k_3} P(A_{k_1}A_{k_2}A_{k_3})\\
&= \binom{4}{1}\frac{\binom{n-2}{2}}{\binom{n-1}{3}} - \binom{4}{2}\frac{\binom{n-3}{1}}{\binom{n-1}{3}} + \binom{4}{3}\frac{\binom{n-4}{0}}{\binom{n-1}{3}}.
\end{aligned}
$$

Here we have use Proposition 4.4 in Chapter 2 (the Inclusion-Exclusion Principle). We don't need to take the intersections of four events here, because that would mean each of the four boxes has just one ball in it, which is impossible because the number of balls is $n \geq 5$.

3. [15=5+5+5 points]
(a) Randomly choose k cards without replacement from a deck of 52 cards. Find the probability that the cards contain no aces.

Solution. To get no aces we must choose k cards from among the 48 non-aces:

$$\text{probability} = \frac{\binom{48}{k}}{\binom{52}{k}}.$$

(b) Instead suppose you choose k cards with replacement (meaning you randomly choose a card, make a note of what it is, then replace it, and repeat). Find the probability that the cards contain no aces.

Solution. To get no aces, each of the k times we choose a card it must be from among the 48 non-aces:

$$\text{probability} = \frac{48^k}{52^k}.$$

(c) For a certain unfair die, we know the probability of rolling the number k equals $\frac{k}{21}$ (for $k = 1, 2, 3, 4, 5, 6$).

First roll the die and then choose k cards with replacement, from the deck. Find the probability that the chosen cards contain no aces.

Solution 1. We condition on the result of rolling the die:

$$\begin{aligned} P(\text{no aces}) &= \sum_{k=1}^{6} P(\text{no aces}|\text{die} = k)P(\text{die} = k) \\ &= \sum_{k=1}^{6} \frac{48^k}{52^k}\frac{k}{21}. \end{aligned}$$

Our conditioning here relies on the mutual exclusivity of the events "die= k", for $k = 1, \ldots, 6$.

4. [10 points] A parallel system functions whenever at least one of its components is working. Consider a parallel system with $n \geq 2$ components and suppose that each component independently works with probability $\frac{2}{3}$.

Assuming the system is functioning, find the probability that the first two components are working.

Solution.

$$
\begin{aligned}
&P(\text{first two components working}|\text{system is functioning}) \\
&= \frac{P(\text{first two components working and system is functioning})}{P(\text{system is functioning})} \\
&= \frac{P(\text{first two components working})}{1 - P(\text{system is not functioning})} \\
&= \frac{\left(\frac{2}{3}\right)^2}{1 - \left(\frac{1}{3}\right)^n}.
\end{aligned}
$$

For the numerator in the second and third lines, notice that if the first two components are working, then automatically the system is functioning.

5. [20 points] People come in 10 types, with a randomly chosen person having probability $\frac{j}{55}$ of being type j, for $j = 1, 2, \ldots, 10$. The probability that a person of type j likes wine is $\frac{j-1}{45}$.

Given that a person likes wine, what is the probability that the person is of type 3?

Solution. We use Bayes' Law:

$$\begin{aligned} P(\text{type 3}|\text{likes wine}) &= \frac{P(\text{type 3 and likes wine})}{P(\text{likes wine})} \\ &= \frac{P(\text{likes wine}|\text{type 3})P(\text{type 3})}{\sum_{k=1}^{10} P(\text{likes wine}|\text{type k})P(\text{type k})} \\ &= \frac{\frac{3-1}{45}\frac{3}{55}}{\sum_{k=1}^{10} \frac{k-1}{45}\frac{k}{55}} \\ &= \frac{1}{55}. \end{aligned}$$

6. [10 points] A hiker is missing and is thought to be in one of two regions: Region 1 with probability $\frac{1}{3}$ and Region 2 with probability $\frac{2}{3}$. The probability that a search of Region 1 will find the hiker is $1 - \beta_1$ if the hiker is in Region 1, and the probability that a search of Region 2 will find the hiker is $1 - \beta_2$ if the hiker is in Region 2.

A search of Region 1 has failed to find the hiker.

(a) Find the probability that the hiker is in Region 1? In Region 2?

Solution. Bayes' Law again:

$$\begin{aligned}
&P(\text{hiker in Region 1}|\text{Region 1 search fails})\\
&= \frac{P(\text{hiker in Region 1 and Region 1 search fails})}{P(\text{Region 1 search fails})}\\
&= P(\text{Region 1 search fails}|\text{hiker in Region 1})P(\text{hiker in Region 1})/\\
&\quad [P(\text{Region 1 search fails}|\text{hiker in Region 1})P(\text{hiker in Region 1})\\
&\quad +P(\text{Region 1 search fails}|\text{hiker in Region 2})P(\text{hiker in Region 2})]\\
&= \frac{\beta_1 \cdot \frac{1}{3}}{\beta_1 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3}} = \frac{\beta_1}{\beta_1 + 2}.
\end{aligned}$$

For Region 2, either argue similarly, or just take 1 minus the previous answer to get

$$P(\text{hiker in Region 2}|\text{Region 1 search fails}) = 1 - \frac{\beta_1}{\beta_1 + 2} = \frac{2}{\beta_1 + 2}.$$

(b) Which region should be searched next, to have the best chance of finding the hiker? (The answer depends on β_1 and β_2.)

Solution. If the next search is in Region 1, then it will find the hiker with probability $(1 - \beta_1)\frac{\beta_1}{\beta_1+2}$, which is obtained by multiplying the chance of finding the hiker if she is in Region 1 by the probability that she is actually in Region 1.

If the next search is in Region 2, then similarly it will find the hiker with probability $(1 - \beta_2)\frac{2}{\beta_1+2}$.

So Region 1 should be searched next if

$$(1 - \beta_1)\frac{\beta_1}{\beta_1 + 2} > (1 - \beta_2)\frac{2}{\beta_1 + 2}.$$

Otherwise Region 2 should be searched next.

NAME:

Math 461 Spring 2005 — Test 2

Total points: **100**. Do all questions. **Explain** all answers. You may use without comment any formulas from class for the density, expectation or variance of random variables. No notes, books, or electronic devices.

1. [20 points] Your average number of driving accidents is 0.2 per year. You take a defensive driving course that has a 70% chance of reducing your rate to 0.1, and a 30% chance of leaving your rate at 0.2.

Given that you have no accidents in the following year, find the probability that your rate is still 0.2.

2. [20=13+7 points] A total of $n \geq 2$ people are to be tested for a certain disease. They will have their blood samples pooled together, so that if the test is negative, one test will suffice for all n people. But if the test is positive, then each of the n people will also be individually tested (so that $n+1$ tests will be done in total). Assume 10% of the people have the disease.

(a) Compute the expected number of tests, and the variance.

(b) A large number N of people are to be tested. Into what group size n should they be divided, in order to minimize the expected total number of tests? (Just explain *how* to find the answer. Actually finding the numerical answer would require a calculator.)

3. [20=7+7+6 points] Suppose $X \sim \text{Normal}(5, 1)$.

(a) Evaluate $P(X < 4)$.

(b) Evaluate $E[X^2]$.

(c) Find the probability that X takes a value more than 2 standard deviations away from its mean.

4. [20 points] A certain baseball player has batting average $p = .2$ (meaning his chance of a hit is .2, in each at-bat). Show that the probability of the player getting fewer than 200 hits in his next 900 at-bats is *approximately* .95.

Hint. $18/12 = 1.5$ and $0.5/12 \approx 0.04$.

Be sure to briefly describe your assumptions, and justify the applicability of your method.

5. [20=8+7+5 points] Buses arrive randomly at an average rate of 10 per hour. Let X equal the length of time (in hours) after 1pm until the third bus arrives.

[If you can't do the following problems for the *third* bus, then do them for the *first* bus, for half-credit.]

(a) What kind of random variable is X? Give a formula for its density.

(b) Evaluate the expected arrival time of the third bus, and find the variance.

(c) Write down an explicit integral for the probability that the arrival time of the third bus lies within one standard deviation of the expected arrival time. (You need not evaluate your integral.)

6. *Extra credit* [10 points] Show that the square of a Standard Normal random variable is a Gamma($\frac{1}{2}, \frac{1}{2}$) random variable.

Math 461 Spring 2005 — Test 2 Solutions

Total points: **100**. **Explain** all answers. You may use without comment any formulas from class for the density, expectation or variance of random variables. No notes, books, or electronic devices.

1. [20 points] Your average number of driving accidents is 0.2 per year. You take a defensive driving course that has a 70% chance of reducing your rate to 0.1, and a 30% chance of leaving your rate at 0.2.

Given that you have no accidents in the following year, find the probability that your rate is still 0.2.

Solution.

$$
\begin{aligned}
&P(\text{rate is } 0.2 \mid \text{no accidents}) \\
&= \frac{P(\text{rate is } 0.2 \text{ and no accidents})}{P(\text{no accidents})} \\
&= \frac{P(\text{no accidents} \mid \text{rate is } 0.2)P(\text{rate is } 0.2)}{P(\text{no accidents} \mid \text{rate is } 0.2)P(\text{rate is } 0.2) + P(\text{no accidents} \mid \text{rate is } 0.1)P(\text{rate is } 0.1)} \\
&= \frac{e^{-0.2}(0.3)}{e^{-0.2}(0.3) + e^{-0.1}(0.7)}
\end{aligned}
$$

where we are assuming X =number of accidents per year is a **Poisson** random variable with rate either $\lambda = 0.2$ or $\lambda = 0.1$. We have used that $P(X = 0) = e^{-\lambda}$ for a Poisson random variable.

The answer can be simplified to $3/(3 + 7e^{0.1}) \simeq 0.28$.

2. [20=13+7 points] A total of $n \geq 2$ people are to be tested for a certain disease. They will have their blood samples pooled together, so that if the test is negative, one test will suffice for all n people. But if the test is positive, then each of the n people will also be individually tested (so that $n+1$ tests will be done in total). Assume 10% of the people have the disease.

(a) Compute the expected number of tests, and the variance.

Solution. Write X for the number of tests. Then

$P(X = 1) = (0.9)^n$, since $X = 1$ happens when no-one has the disease, and so $P(X = n+1) = 1 - (0.9)^n$.

Hence

$$\begin{aligned}
E[X] &= 1 \cdot P(X = 1) + (n+1) \cdot P(X = n+1) \\
&= (0.9)^n + (n+1)[1 - (0.9)^n] \\
Var(X) &= E[X^2] - (E[X])^2 \\
&= 1^2 \cdot P(X = 1) + (n+1)^2 \cdot P(X = n+1) - (E[X])^2 \\
&= (0.9)^n + (n+1)^2[1 - (0.9)^n] - ((0.9)^n + (n+1)[1 - (0.9)^n])^2 \\
&= (0.9)^n[1 - (0.9)^n]n^2
\end{aligned}$$

(b) A large number N of people are to be tested. Into what group size n should they be divided, in order to minimize the expected total number of tests? (Just explain *how* to find the answer. Actually finding the numerical answer would require a calculator.)

Solution. The N people can be divided into approximately N/n groups of size n. So the expected number of tests is

$$\frac{N}{n}((0.9)^n + (n+1)[1 - (0.9)^n]),$$

by the expectation formula in part (a).

We want to minimize this expression over all possible choices of $n \in \{2, 3, \ldots, N\}$ (with N being fixed). The easiest way to find the minimum would be to have a computer evaluate the expression for all n. Or we could regard n as a real number (rather than an integer) and try to find the minimum using methods of calculus.

3. [20=7+7+6 points] Suppose $X \sim \text{Normal}(5, 1)$.

(a) Evaluate $P(X < 4)$. *Solution.* Write $Z = (X - \mu)/\sigma$ for the standard **normal** random variable, with mean zero and standard deviation 1. Then

$$\begin{aligned}
P(X < 4) &= P\left(\frac{X-\mu}{\sigma} < \frac{4-\mu}{\sigma}\right) \\
&= P\left(Z < \frac{4-5}{1}\right) \\
&= P(Z < -1) \\
&= \Phi(-1) \\
&= 1 - \Phi(1) \\
&\simeq 1 - 0.8413 \\
&= 0.1587
\end{aligned}$$

(b) Evaluate $E[X^2]$. *Solution.*

$$\begin{aligned}
E[X^2] &= Var(X) + (E[X])^2 \\
&= 1 + 5^2 = 26
\end{aligned}$$

(c) Find the probability that X takes a value more than 2 standard deviations away from its mean. *Solution.*

$$\begin{aligned}
P(|X - \mu| > 2\sigma) &= P(X < \mu - 2\sigma) + P(X > \mu + 2\sigma) \\
&= 2P(X > \mu + 2\sigma) \qquad \text{by symmetry} \\
&= 2P\left(\frac{X-\mu}{\sigma} > 2\right) \\
&= 2P(Z > 2) \\
&= 2(1 - P(Z \leq 2)) \\
&= 2(1 - \Phi(2)) \\
&= 2(1 - 0.9772) \simeq 0.0456
\end{aligned}$$

4. [20 points] A certain baseball player has batting average $p = .2$ (meaning his chance of a hit is .2, in each at-bat). Show that the probability of the player getting fewer than 200 hits in his next 900 at-bats is *approximately* .95.

Hint. $18/12 = 1.5$ and $0.5/12 \approx 0.04$.

Be sure to briefly describe your assumptions, and justify the applicability of your method.

Solution. Write $X =$ #hits in the next 900 at-bats. Then

$$X \sim \textbf{Binomial}(n, p)$$

with $n = 900, p = 0.2$. *(We assume the at-bats are independent, so that we can regard them as Bernoulli trials.)*

The mean and variance of X are

$$\mu = E[X] = np = 180$$
$$\sigma^2 = Var(X) = np(1-p) = 144$$

so that $\sigma = 12$.

We use the **Normal** Approximation to the Binomial *(which is valid since n is large and $\sigma^2 \geq 10$)*: the probability of fewer than 200 hits in the next 900 at-bats is

$$\begin{aligned} P(X < 200) &= P(X \leq 199.5) \qquad \text{using the “continuity correction”} \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{199.5-180}{12}\right) \\ &= P(Z \leq 19.5/12) \\ &= \Phi(1.62) \\ &\simeq 0.95 \end{aligned}$$

Remark. We use the **Poisson** approximation to the Binomial when n is large and $\lambda = np$ is moderate. (In this problem, $np = 180$ is too large to be called “moderate”.) We use the **Normal** approximation to the Binomial when n is large and $\sigma^2 = np(1-p) \geq 10$.

5. [20=8+7+5 points] Buses arrive randomly at an average rate of 10 per hour. Let X equal the length of time (in hours) after 1pm until the third bus arrives.

[If you can't do the following problems for the *third* bus, then do them for the *first* bus, for half-credit.]

(a) What kind of random variable is X? Give a formula for its density.

Solution. X is the waiting time till the third event, and so $X \sim \mathbf{Gamma}(3, \lambda)$ where the rate is $\lambda = 10$ buses per hour. The Gamma density with $\alpha = 3$ and $\lambda = 10$ is

$$f(x) = \begin{cases} \frac{10e^{-10x}(10x)^{3-1}}{\Gamma(3)} & x > 0, \\ & x \leq 0. \end{cases}$$

Remember $\Gamma(3) = 2! = 2$.

(b) Evaluate the expected arrival time of the third bus, and find the variance.

Solution.

$$E[X] = \frac{\alpha}{\lambda} = \frac{3}{10}$$
$$Var(X) = \frac{\alpha}{\lambda^2} = \frac{3}{100}$$

Since 3/10 of an hour equals 18 minutes, we conclude the expected arrival time is 1:18pm.

(c) Write down an explicit integral for the probability that the arrival time of the third bus lies within one standard deviation of the expected arrival time. (You need not evaluate your integral.)

Solution.

$$P(|X - \mu| < \sigma) = P(|X - 3/10| < \sqrt{3}/10)$$
$$= \int_{(3-\sqrt{3})/10}^{(3+\sqrt{3})/10} \frac{10e^{-10x}(10x)^{3-1}}{2}\, dx$$

6. *Extra credit* [10 points] Show that the square of a Standard Normal random variable is a Gamma$(\frac{1}{2}, \frac{1}{2})$ random variable.

Solution. Let $X \sim \text{Normal}(0,1)$ and $Y = X^2$. Then for $0 \le a < b$ we have

$$\begin{aligned}
P(a \le Y \le b) &= P(\sqrt{a} \le |X| \le \sqrt{b}) \\
&= 2P(\sqrt{a} \le X \le \sqrt{b}) \qquad \text{by symmetry of the bell curve} \\
&= 2\int_{\sqrt{a}}^{\sqrt{b}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \qquad \text{by using the Standard Normal density} \\
&= 2\int_{a}^{b} \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2y^{1/2}}\, dy \qquad \text{by letting } x = y^{1/2} \\
&= \int_{a}^{b} \frac{\frac{1}{2} e^{-(1/2)y} (\frac{1}{2}y)^{1/2-1}}{\Gamma(1/2)}\, dy \qquad \text{since } \Gamma(1/2) = \sqrt{\pi}.
\end{aligned}$$

Thus the density of Y is precisely the Gamma$(\frac{1}{2}, \frac{1}{2})$ density.

Remark. This problem has nothing to do with squaring the density function!

NAME:

Math 461 Spring 2005 — Final Exam

Total points: **200**. Do all questions. **Explain** all answers. You may use without comment any formulas from class for the density, expectation or variance of random variables. No notes, books, or electronic devices.

1. [15 points] Fifteen distinct balls are randomly divided among players E, F and G, with each player getting five balls.

If ten of the balls are White and five are Black, then find the probability player E gets 2 White balls, player F gets 3 White balls, and player G gets 5 White balls.

2. [20 points] If a 5-card poker hand contains at least 2 aces, then find the probability that it is a full house (meaning it has the form xxxyy).

3. [15 points] Suppose $X \sim \text{Exponential}(\lambda)$. Show the density of $Y = \log X$ is $g(y) = \lambda \exp[y - \lambda e^y]$.

4. [25=17+8 points] An ambulance travels back and forth, at a constant speed, along a straight road of length L. At a certain moment of time an accident occurs at a point X uniformly distributed on the road. Assume that the ambulance's location Y at the moment of the accident is also uniformly distributed, and is independent of the location of the accident.

(a) Show that the distribution function of the distance between accident and ambulance (along the road) is $\frac{a}{L}(2 - \frac{a}{L})$. Also: for which a-values is this answer valid?

(b) Suppose now the road is circular rather than straight (but still with length L). Find the distribution function of the distance between accident and ambulance (along the road).

5. [20=14+6 points] Let $X \sim \text{Exponential}(\lambda)$ and $Y \sim \text{Exponential}(\mu)$ be independent exponential random variables.

(a) Find the distribution of $Z = Y/X$.

(b) Compute $P(X < Y)$.

6. [20=10+10 points] "Big" apples have weight approximately normally distributed with mean 160 grams and standard deviation 15. "Large" apples have weight approximately normally distributed with mean 170 grams and standard deviation 20.

Pick one of each type of apple, independently.

(a) Find the probability that the two apples together weigh more than 350 grams. Explain.

(b) Find the probability that the "big" apple weighs more than the "large" one. Explain.

7. [15=10+5 points] [**Do #7 or #8 but not both.**] Suppose $X, Y, Z \sim$ Geometric(p) are independent random variables.

Evaluate $P(\min(X, Y, Z) \leq k)$.

(b) Interpret part (a) in terms of coin tossing.

8. [15=7+8 points] [**Do #7 or #8 but not both.**] There are 3 storms, on average, in a good winter. In a bad winter there are 5 storms, on average. Next winter will be good with probability 0.4 and bad with probability 0.6.

(a) Show the expected number of storms next winter is 4.2. Explain any assumptions you make.

(b) Show the variance of the number of storms next winter is 5.16.

9. [35=3+12+5+15 points] Choose a point (X, Y) randomly according to the uniform distribution in the square $\{(x, y) : -1 \leq x \leq 1, -1 \leq y \leq 1\}$.

(a) Write down the joint density $f(x, y) =$

(b) Define a new random variable

$$R = \begin{cases} \sqrt{X^2 + Y^2} & \text{if } \sqrt{X^2 + Y^2} < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Show $E[R] = \pi/6$. Then find $Var(R)$.

(c) Let $R_1, R_2, \ldots$ be independent random variables of the type in part (b), corresponding to independent choices of points $(X_1, Y_1), (X_2, Y_2), \ldots$ in the square.

Explain what the Central Limit Theorem says about the random variable $R_1 + R_2 + \cdots + R_n$.

(d) Write $\overline{R}_n = (R_1 + \cdots + R_n)/n$ for the sample mean. Find an estimated value of n for which $P(|\overline{R}_n - \frac{\pi}{6}| < 0.01) \leq 0.95$.

You may leave your answer in terms of the standard deviation σ found in part (b).

10. [35=10+10+15 points] A group of 20 gloves, consisting of 8 left gloves and 12 right gloves, is randomly arranged into 10 pairs of 2 gloves each. Let

$$X = \text{the number of pairs that consist of a Left and a Right.}$$

(a) Show $E[X] = 10p$ where $p = 48/95$.

(b) Find $Var(X)$.

(c) Compute the probability mass function of X.
Hint. For partial credit, try finding $P(X = 0)$ and $P(X = 1)$.