1.) One hundred and one passengers bought tickets on a 101-seat carriage. One seat was reserved for each passenger. The first 100 passengers took the seats at random so that all 101! possible seating arrangements (with one empty seat) are equally likely. The last passenger insisted on taking the assigned seat. If that seat is occupied, then the passenger in that seat has to move to the corresponding assigned seat, and so on. Compute the expected value of the number M of passengers who have to change their seats. [HINTS: one method is to use a recursion in n, for n in the role of 101, for the expectation, without knowing the distribution of M. Another method is to find the distribution of M explicitly.]

2.) Suppose \( P(X = k) = p_k \) and \( p_1 + p_2 + \cdots = 1 \). Suppose that \( X, X_1, X_2, \ldots, X_n \) are independent and identically distributed. Let \( S = \sum_{1 \leq i < j \leq n} 1(X_i = X_j) \) be the number of matching (unordered) pairs, and let \( T = \sum_{1 \leq i < j < k \leq n} 1(X_i = X_j = X_k) \) be the number of matching (unordered) trios. For \( r = 1, 2, 3, \ldots, \), let \( f_r = \sum p_r^i \), so that \( f_1 = 1 \).
   a) Give a simple expression for \( E S \) in terms of \( n, f_2 \).
   b) Give a simple expression for \( E T \) in terms of \( n, f_3 \).
   c) Give a simple expression for \( \text{Var} S \) in terms of \( n, f_2, f_3, f_4 \).

3.) Let \( X, X_1, X_2, X_3, X_4 \) be independent standard exponentially distributed random variables, so that \( P(X > x) = e^{-x} \) for \( x > 0 \). Write \( S_n = X_1 + \cdots + X_n \). The goal is to show that the triple \( (S_1/S_4, S_2/S_4, S_3/S_4) \) is distributed like the order statistics of three independent standard uniform \((0, 1)\) random variables.
   a) Give a reason why the density of \( S_4 \) is \( f(t) = t^3 e^{-t}/6 \) for \( t > 0 \). You may either quote the density for the Gamma family in general, or you may argue about the time of the fourth arrival in a standard, rate 1 Poisson process, or you may carry out the four-fold convolution!
   b) With \((U_1, U_2, U_3)\) distributed uniformly in the cube \((0,1)^2\), and \( U_{[i]} \) defined to be the \( i \)th smallest of \( U_1, U_2, U_3 \), show why the density of \((U_{[1]}, U_{[2]}, U_{[3]})\) is \( g(x, y, z) = 6 \), on the region \( 0 < x < y < z < 1 \).
   c) Show, with detail, why the triple \((S_1/S_4, S_2/S_4, S_3/S_4)\) is distributed like the order statistics of three independent standard uniform \((0,1)\) random variables, AND that this triple is independent of \( S_4 \). This should include calculation of both a 4 by 4 Jacobian matrix, and calculation of its determinant, also known as the “Jacobian”.
   d) Consider three independent uniform \((0, 1)\) variables. Compute the probability that the largest exceeds the sum of the other two.
In the Polya Urn model, $w \geq 1$ white balls and $b \geq 1$ black balls are placed in an urn at time $0$, and at times $1, 2, \ldots$ a ball is chosen uniformly from the urn independent of the past, and replaced back into the urn with one additional ball of the same color.

a. A vector $(X_1, \ldots, X_n)$ of random variables is said to be exchangeable

$$(X_1, \ldots, X_n) =_d (X_{\pi(1)}, \ldots, X_{\pi(n)})$$

for all permutations $\pi$, where $=_d$ denotes equality in distribution. If $X_i$ is the indicator that a white ball is drawn from the urn at time $i$, prove that $(X_1, \ldots, X_n)$ is exchangeable.

b. Find the mean and variance of $S_n = X_1 + \cdots + X_n$, the total number of white balls added to the urn up to time $n$.

With $a$ and $b$ positive numbers, a needle of length $l \in (0, \min(a, b)]$ is dropped randomly on a rectangular grid consisting of an infinite number of parallel lines distance $a$ apart, and, perpendicular to these, an infinite number of parallel lines distance $b$ apart. Let $A$ and $B$, respectively, be the events that the needle intersects the group of lines at distance $a$ and $b$ apart.

a. Show $P(A) = \frac{2l}{\pi a}$ and $P(B) = \frac{2l}{\pi b}$. Hint: The angle $\theta$ giving the orientation of the needle might be taken as uniform from $[0, 2\pi)$, but by symmetry, one may assume that the angle is uniformly taken from $[0, \pi/2)$.

b. Determine $P(A \cap B)$ and verify that $A$ and $B$ are strictly negatively correlated, that is, that $P(A \cap B) < P(A)P(B)$.

A total of $k$ boys and $n - k$ girls sit around a circular table, with all $n!$ arrangements equally likely. Compute the mean and variance of the number $Y$ of pairs of boy/girl neighbors. Note: In the circular arrangement GBGGBB, since the first $G$ and last $B$ are neighbors, $Y = 4$. 
Math 505a 2013 Spring Qualifying Exam

1. a) Let $X$ and $Y$ be square integrable random variables such that

$$E(X|Y) = Y \quad \text{and} \quad E(Y|X) = X.$$  \hspace{1cm} (1)

Show that

$$P(X = Y) = 1.$$  \hspace{1cm} (2)

b) Prove that (1) implies (2) under the weakened assumption that $X$ and $Y$ are integrable.

2. Suppose $k$ balls are tossed into $n$ boxes, with all $n^k$ possibilities equally likely. Let $D$ be the number of boxes that contain exactly 2 balls.

a) Compute $p := P($ exactly 2 balls land in box 1$)$.

b) In terms of $p$, give an exact expression for the mean $ED$.

c) Compute $r := P($ exactly 2 balls land in box 1 and exactly 2 balls land in box 2$)$.

d) Give an exact expression for the second moment $ED^2$ in terms of $p$ and $r$.

e) Compute the variance of $D$.

3. a) Suppose $g(u) := Eu^S$ is the probability generating function of a non-negative integer valued random variable $S$ satisfying $P(S > 0) > 0$. Let $T$ be distributed as $S$, conditional on the event $S > 0$. Express $h(u) := Eu^T$, the probability generating function of $T$, in terms of $g(u)$.

In parts b) and c) below, $N$ is a nonnegative integer valued random variable with probability generating function $f(u) := Eu^N$, and $S$ is the number of heads in $N$ tosses of a $p \in (0, 1)$ coin, with all coin tosses having probability $p$ of coming up heads, independently of each other and of $N$. 
b) Write the probability generating function \( g(u) := E u^S \) of \( S \) in a simple form.

c) Now combine parts a) and b): what is the probability generating function \( h \) of the number \( T \) of heads, in \( N \) tosses of a \( p \)-coin, conditional on getting at least one head, when \( N \) has probability generating function \( f \)? Parts d,e) can be worked on even if you are stumped by a,b,c).

d) Suppose someone claims that for \( \alpha \in (0, 1) \), the function

\[
    f(u) := 1 - (1 - u)^\alpha
\]

is a probability generating function of a nonnegative, non constant integer valued random variable \( N \). What properties of \( f \) must you check? Is the hypothesis \( \alpha > 0 \) used? What happens in the cases \( \alpha = 0, \alpha = 1 \) and \( \alpha > 1 \)?

e) Combine parts a)-d), that is suppose \( \alpha \in (0, 1) \), \( N \) has the generating function \( f(u) := 1 - (1 - u)^\alpha \), and \( T \) is the number of heads in \( N \) tosses of a \( p \)-coin, conditional on getting at least one head. Do \( N \) and \( T \) have the same distribution?
1.) In an infinite sequence of independent trials, events $A, B$ are mutually exclusive, with $a = \mathbb{P}(A) > 0$ and $b = \mathbb{P}(B) > 0$.

a.) What is the probability that $A$ will occur before $B$?

b.) In repeated independent tosses of a pair of fair dice, what is the probability that the sum 3 will occur before the sum 7?

2.) Let $X$ and $Y$ be independent, standard normal. Let $W = X + Y$ and $Z = X - Y$.

a.) Show that $W$ and $Z$ are independent.

b.) Simplify $\mathbb{E}(X + 2Y | Z)$.

c.) Simplify $\mathbb{E}(X | X > 0)$.

3.) $n$ balls are placed into $d$ boxes at random, with all $d^n$ possibilities equally likely. Assume $d > 8$. Let $X$ be the number of empty boxes.

a) Calculate and simplify: $\mathbb{E}X = \ldots$

b) Calculate and simplify: $\text{Var} X = \ldots$

c) Let $A$ be the event that boxes 1,2,3,4 are all empty, $B$ be the event that boxes 3,4,5,6 are all empty, and $C$ be the event that boxes 5,6,7,8 are all empty. Compute exactly, $\mathbb{P}(A \cup B \cup C) = \ldots$

d) Let $D$ be the event that no box receives more than 1 ball. Fix $a \in (0,1)$. If both $n,d \to \infty$ together, what relation must they satisfy in order to have $\mathbb{P}(D) \to a$?
1. A person plays a sequence of $m$ games. He wins the $n$th game with probability $a_n$ independently of the other games. Every time that he wins two consecutive games, he is rewarded with $\$1$; let $R$ be the total reward, so that $0 \leq R \leq m - 1$.
   (i) As a function of $a_1, a_2, \ldots, a_m$, give exact expressions for $E(R)$ and $\text{Var}(R)$.
   (ii) Now suppose that $m = 100$ and $a_n = .1$ for all $n$. Simplify numerically $E(R)$ and $\text{Var}(R)$.

2. Assume that $X_1, X_2, \cdots$ are independent and identically distributed, each with density
   \[ f(x) = x/2 \text{ for } 0 < x < 2. \]
   For each of the following random variables, simplify the density or cumulative distribution function; you may choose either one, for each random variable.
   a) $S = X_1 + X_2$.
   b) $L = \min(X_1, \ldots, X_{100})$.
   c) $R = X_1/X_2$.
   d) $M$ = the $10^{th}$ smallest of $X_1, \ldots, X_{100}$.

3. Let $X_1, X_2, \cdots$ be uncorrelated random variables with $E[X_i] = \mu$ and $var(X_i) \leq C < \infty$. If $S_n = X_1 + \cdots + X_n$, show that as $n \to \infty$, $S_n/n \to \mu$ in probability. That is, prove that for any $\varepsilon > 0$,
   \[ \lim_{n} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0. \]
   Even if you are not comfortable with limits, simply give the best upper bound that you can, of the form
   \[ P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \underline{\text{______}}. \]
Answer all three questions. Partial credit will be awarded, but in the event that you can
not fully solve a problem you should state clearly what it is you have done and what you
have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your
score. Start each problem on a fresh sheet of paper, and write on only one side of the paper.
If you find that a calculation leads to something impossible, such as a negative probability
or variance, indicate that something is wrong, but show your work anyway.

1. Let $A$ and $B$ be two events with $0 < P(A) < 1, 0 < P(B) < 1$. Define the random
variables $\xi = \xi(\omega)$ and $\eta = \eta(\omega)$ by

$$\xi(\omega) = \begin{cases} 5 & \text{if } \omega \in A; \\ -7 & \text{if } \omega \notin A; \end{cases} \quad \eta(\omega) = \begin{cases} 2 & \text{if } \omega \in B; \\ 3 & \text{if } \omega \notin B. \end{cases}$$

True or false: the events $A$ and $B$ are independent if and only if the random variables $\xi$ and
$\eta$ are uncorrelated? If you think this is true, then provide a proof. If you think this is false,
then give a counter-example.

2. $n$ people each roll one fair die. For each (unordered) pair of people that get the
same number of spots, that number of spots is scored, with $S$ for the total score achieved
among the $\binom{n}{2}$ pairs of people. For example, if there are $n = 10$ people, and they roll
$1,2,2,3,4,4,4,6$ then $S = 2 + 2 + 4 + 4 + 4 + 4 + 4 + 6$ since there are three pairs of
people matching 2 and six $\binom{n}{2}$ pairs of people scoring 4.

(a) Simplify $E S$.
(b) Simplify $E S^2$.
[HINT: Consider $S$ as the sum of $\binom{n}{2}$ random variables $S_{i,j}$, where $S_{i,j}$ is $k$ if persons $i$
and $j$ both roll $k$, and zero otherwise.]

3. Let $X$ be a standard normal random variable and, for $a > 0$, define the random
variable $Y_a$ by

$$Y_a = \begin{cases} X, & \text{if } |X| < a, \\ -X, & \text{if } |X| > a. \end{cases}$$

(a) Verify that $Y_a$ is a standard normal random variable.
(b) Express $\rho(a) = E(XY_a)$ in terms of the probability density function $\varphi = \varphi(x)$ of $X$.
(c) Is there a value of $a$ for which $\rho(a) = 0$?
(d) Does the pair $(X, Y_a)$ have a bivariate normal distribution? Explain your reasoning.
Answer as many questions as you can. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. If you cannot do part (a) of a problem, you can still get credit for (b), (c) etc. by assuming the answer to (a). Start each problem on a fresh sheet of paper, and write on only one side of the paper.

(1) (a) There are \( n = 100 \) balls, labelled 1 to \( n \), and these are thrown into \( n \) boxes, also labelled 1 to \( n \); all \( n^n \) outcomes are equally likely. Whenever ball \( i \) lands in box \( j \), and \( |i - j| \leq 1 \), a point is scored, so the total score, call it \( X \), can take on any value from 0 to \( n \). Note: ball 1 in box \( n \) scores nothing, and ball \( n \) in box 1 scores nothing. Compute exactly, and simplify, as either an expression in \( n \), or a decimal: \( E(X) \) and \( \text{Var}(X) \).

(b) Pick the closest approximation: in part (a), \( P(X = 0) \) is close to 1, 1/3, 1/20, 1/100. You may reason informally, but you must describe your reasoning, which should involve the approximate distribution of \( X \).

(c) Change the story in (a) to: there are \( n = 100 \) cards, numbered 1 to \( n \), and they are placed in positions 1 to \( n \) around a circle at random, so that all \( n! \) outcomes are equally likely. Whenever card \( i \) is placed in position \( i \) or \( i + 1 \), a point is scored, so the total score, call it \( Y \), can take on any value from 0 to \( n \). Here for card \( n \), since the cards are in a circle, “position \( n + 1 \)” should be interpreted as position 1. Compute exactly, and simplify, as either an expression in \( n \), or a decimal: \( E(Y) \) and \( \text{Var}(Y) \).

(2) (a) A coin comes up heads with probability \( p \in [0, 1] \); it is tossed independently \( n \) times, and \( X \) is the total number of heads. Simplify the generating function

\[
G(s) = E(s^X),
\]

and show how derivatives of \( G \) can be used to calculate \( E(X) \) and \( E(X^2) \).

(b) Random variables \( U, U_1, U_2, \ldots, U_n \) are independent, and uniformly distributed in \( [0,1] \). Let

\[
Y = \sum_{i=1}^{n} 1_{\{U_i < U\}}
\]

be the sum of indicators, counting how many of the \( U_i \) are less than \( U \). Simplify the generating function

\[
H(s) = E(s^Y),
\]

and then simplify the ratio, \( P(Y = 2)/P(Y = 1) \). HINT: conditionally on \( U = p \), this \( Y \) is distributed as the \( X \) in part (a).
(3) Let $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ be iid uniform on $(0, 1)$.
(a) Compute the probability density function of $\max(Y_1, \ldots, Y_m)$.
(b) Compute the probability density function of

$$\frac{\max(X_1, \ldots, X_n)}{\max(Y_1, \ldots, Y_m)}.$$ 

HINT: One method (not the only one) is to use conditioning.
Answer as many questions as you can. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. If you cannot do part (a) of a problem, you can still get credit for (b), (c) etc. by assuming the answer to (a). Start each problem on a fresh sheet of paper, and write on only one side of the paper.

(1) You have a choice to roll a fair die either 100 times or 1000 times. For each of the following outcomes, state whether it is more likely with 100 rolls, or with 1000 rolls. Justify your answer, but you do not need to give a full formal proof.
   (a) The number 1 shows on the die between 15% and 20% of the time.
   (b) The number showing is at most 3, at least half the time.
   (c) The number showing is 2 or 5, at least half the time.

(2) Let $X$ and $Y$ be independent exponential random variables with parameters $\lambda$ and $\mu$ (that is, $E(X) = 1/\lambda$ and $E(Y) = 1/\mu$), and let $Z = \min(X, Y)$.
   (a) Show that $Z$ is independent of the event $X < Y$. In other words, show the event $Z \leq t$ is independent of $X < Y$ for all $t$.
   (b) Find the distribution of $\max(X - Y, 0)$.

(3) Let $X_1, X_2, \ldots$ be iid with characteristic function $\phi$. Let $N$ be independent of the $X_i$’s with $P(N = n) = 2^{-n}$ for all $n \geq 1$. Let $Y = \sum_{i=1}^{N} X_i$. Find the characteristic function of $Y$.

(4) $n \geq 4$ men, among whom are Alfred, Bill, Charles and David, stand in a row. Assume that all possible orderings of the $n$ men are equally likely.
   (a) Find the probability that Charles stands somewhere between Alfred and Bill. (Note this does not mean they are necessarily adjacent—there might be other people between Alfred and Bill.)
   (b) Find the probability that David stands somewhere between Alfred and Bill given that Charles stands somewhere between Alfred and Bill.
   (c) Find the expected value and variance of the number of men out of $n$ who stand between Alfred and Bill. (Note Alfred and Bill themselves are not counted in this number.)
MATH 505a QUALIFYING EXAM Monday, February 9, 2015. One hour and 50 minutes, starting at 5pm.

Answer all three questions. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, and write on only one side of the paper. If you find that a calculation leads to something impossible, such as a negative probability or variance, indicate that something is wrong, but show your work anyway.

1. Let $X_n$, $n \geq 1$, be independent random variables such that each $X_n$ has Poisson distribution with mean $\lambda_n$. Prove that if $\sum_{n \geq 1} \lambda_n = +\infty$, then

$$\lim_{n \to \infty} \frac{\sum_{k=1}^{n} X_k}{\sum_{k=1}^{n} \lambda_k} = 1$$

in probability.

2. A deck of cards is shuffled thoroughly. Someone goes through all 52 cards, scoring 1 each time 2 cards of the same value are consecutive. For example 9H,8H,7D,6C,7S,7H,7C, scores 2, once due to 7 of spades next to 7 of hearts, and once more 7 of hearts next to 7 of clubs. Write $X$ for the total score.

a) Compute $\mathbb{E}X$.

b) Compute $\textrm{Var}X$.

c) Compute $\mathbb{P}(X = 39)$.

d) In the line below, circle the number that you think is the closest to the value $\mathbb{P}(X = 0)$ and briefly explain your choice.

\[
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1000' & 500' & 100' & 50' & 20' & 10' & 5' & 2'.
\end{array}
\]

3. Let $S_0, S_1, S_2, \ldots$ be a simple symmetric random walk, i.e. $\mathbb{P}(S_i - S_{i-1} = 1) = \mathbb{P}(S_i - S_{i-1} = -1) = 1/2$, with independent increments. Let $T = \min\{n > 0 : S_n = 0\}$ be the hitting time to zero. Write $\mathbb{P}_a$ for probabilities for the walk starting with $S_0 = a$.

a) What does the reflection principle say about $\mathbb{P}_a(S_n = i, T \leq n)$, for $a > 0$, and $i, n \geq 0$?

b) What does the reflection principle say about $\mathbb{P}_a(S_n \geq i, T > n)$, for $a > 0$, and $i, n \geq 0$? [Hint: telescoping series]

c) For fixed $a > 0$, give asymptotics for $\mathbb{P}_a(T > n)$ as $n \to \infty$. [HINT: Stirling’s formula is that $n! \sim \sqrt{2\pi n} \left(n/e\right)^n$.]

d) Simplify, for fixed $a > 0$,

$$\lim_{n \to \infty} \frac{\mathbb{P}_{a+1}(T > n)}{\mathbb{P}_a(T > n)}.$$
MATH 505a GRADUATE EXAM
Spring 2016

Answer as many questions as you can. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. If you cannot do part (a) of a problem, you can still get credit for (b), (c) etc. by assuming the answer to (a). Start each problem on a fresh sheet of paper, and write on only one side of the paper.

(1) A stick of length 1 is broken at a point uniformly distributed over its length.
   (a) Find the mean and variance of the sum $S$ of the squares of the lengths of the two pieces.
   (b) Find the density function of the product $M$ of the lengths of the two pieces. Note that $M \in [0, \frac{1}{4}]$.

(2) There are two types of batteries in a bin. The life span of type $i$ is an exponential random variable with mean $\mu_i$, $i = 1, 2$. The probability of type $i$ battery to be chosen is $p_i$, with $p_1 + p_2 = 1$. Suppose a randomly chosen battery is still operating after $t$ hours. What is the probability that it will still be operating after an additional $s$ hours?

(3) Fix positive integers $m \leq n$ with $n > 4$. Suppose $m$ people sit at a circular table with $n$ seats, with all $\binom{n}{m}$ seatings equally likely. A seat is called isolated if it is occupied and both adjacent seats are vacant. Find the mean and variance of the number of isolated seats.
MATH 505a PROBABILITY GRADUATE EXAM  
Spring 2017

Answer as many questions as you can. Partial credit will be awarded, but in the event that you can not fully solve a problem you should state clearly what it is you have done and what you have left out. Unacknowledged omissions, incorrect reasoning and guesswork will lower your score. Start each problem on a fresh sheet of paper, and write on only one side of the paper.

(1) Three points are chosen independently and uniformly inside the unit square in the plane. Find the expected area of the smallest closed rectangle that has sides parallel to the coordinate axes and that contains the three points. HINT: Consider what happens with just one coordinate.

(2) Suppose \((X,Y)\) has joint density of the form \(f(x,y) = g(\sqrt{x^2 + y^2})\) for \((x,y) \in \mathbb{R}^2\), for some function \(g\). Show that \(Z = Y/X\) has the Cauchy density \(h(t) = 1/(\pi(1 + t^2))\), \(t \in \mathbb{R}\). HINT: Polar coordinates.

(3) Assume \(\sqrt{3} < C < 2\). Consider a sequence \(X_1, X_2, X_3, \ldots\) of random variables where \(X_1\) is uniform on \([0, 1]\), and where the conditional distribution of \(X_{n+1}\) given \(X_n\) is uniform on \([0, CX_n]\).
   (a) Find the conditional expectation of \((X_{n+1})^r\) given \(X_n\), for \(r \geq 1\).
   (b) Show that \(X_n\) converges to 0 in mean but not in mean square.
   (c) Show that \(X_n\) converges to 0 almost surely.

(4) Suppose that \(n\) boys and \(m\) girls are arranged in a row, and assume that all possible orderings of the \(n + m\) children are equally likely.
   (a) Find the probability that all \(n\) boys appear in a single block.
   (b) Find the probability that no two boys are next to each other.
   (c) Find the expected number of boys who have a girl next to them on both sides.