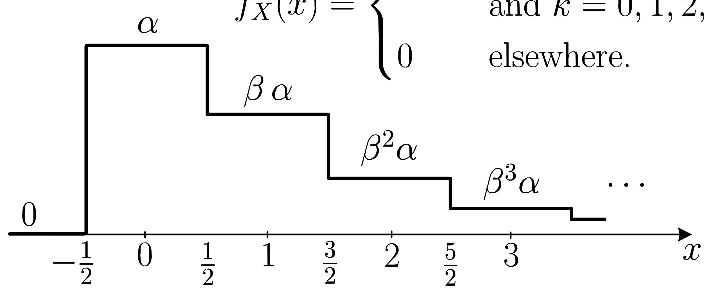
EECS 70 Discrete Math and Probability
Spring 2020 UC Berkeley

Midterm 3

1. Stepping Stones (22 Points)

Consider a continuous random variable X whose PDF is illustrated in the figure below, where $0 < \alpha$ and $0 < \beta < 1$.

$$f_X(x) = \begin{cases} \beta^k \alpha & \text{for } x \in [k - 1/2, k + 1/2), \\ & \text{and } k = 0, 1, 2, 3, \dots \\ 0 & \text{elsewhere.} \end{cases}$$



In one or more parts of this problem, you may or may not find it useful to know that for any $\gamma \in \mathbb{R}$ such that $|\gamma| < 1$, the following identities hold:

$$\sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma} \quad \text{and} \quad \sum_{k=0}^{\infty} k\gamma^k = \frac{\gamma}{(1-\gamma)^2}$$

(a) (3 pts) **Determine** $\Pr(X \geq 0)$**. Justify your answer for full credit.** Express your answer in terms of in terms of α, β, or both.

(HINT: What is $\Pr(X \geq 0)$ *in terms of* $\Pr(X < 0)$*?)*

(b) (5 pts) **Determine α in terms of β. Justify your answer for full credit.**

(HINT: What is the relevant condition that any valid PDF must satisfy?)

(c) (8 pts) Consider a new random variable Y that represents rounding X to the nearest integer. That is, $Y = k$ whenever $X \in [k - \frac{1}{2}, k + \frac{1}{2})$. **Determine** $\mathbb{E}[Y]$**. Show all your work for full credit.** If you're unsure of your expression for α in the previous part, you may leave your expression for the mean in terms of α and β.

(HINT: For natural numbers k, what is $\Pr\big(X \in [k - \frac{1}{2}, k + \frac{1}{2})\big)$*? This can help you make progress towards being able to use the identities given above.)*

(d) (6 pts) **Show that**:

$$P(X \geq b) \leq \frac{\mathbb{E}[X] + \frac{1}{2}}{b + \frac{1}{2}}, \quad \text{for all} \quad b > -\frac{1}{2}.$$

(HINT: Let $c = b + \frac{1}{2}$ and rewrite what you want to show in terms of c. Does this remind you of anything? Can $Z = X + \frac{1}{2}$ ever be negative?)

© UCB EECS 70, Spring 2020. All Rights Reserved. This may not be publicly shared without explicit permission.

2. Disks on a Chess Board (12 pts)

Let n be a positive integer. Say we have $4n$ disks. We want to place them each on a chess board of size $2n \times 2n$ tiles, one at a time. Assume that each tile may host any number of disks, and that when placing each disk, all $(2n)^2$ tiles are equally likely to be selected for disk placement. Different disks are placed independently — say by rolling a $(2n)^2$-sided fair die.

Let the random variable X denote the total number of disks that are placed on the first row of tiles after all disks have been placed.

Let the random variable Y denote the total number of disks that are placed on the second row of tiles after all disks have been placed.

Because there is nothing different about the first row and the second row, X and Y are identically distributed.

Determine $\mathbb{E}[X]$ and Var(X) in terms of n.

However, X and Y are not independent since the number of disks in the first row clearly constrains the number of disks that can be in the second row.

Determine the joint distribution of X and Y, i.e. $P[X = x, Y = y]$ for all $x, y \in \{0, 1, \ldots, 4n\}$. Show all your work for full credit.

(HINT: It might be helpful to write X as a sum of appropriate indicator random variables for the first part of this.)

© UCB EECS 70, Spring 2020. All Rights Reserved. This may not be publicly shared without explicit permission.

3. Waiting in Line at Disneyland (19 pts)

Rohan, Sally, and Tom go to Disneyland and each choose to wait in line for different rides. Each person joins their respective line at the same time.

- Rohan chooses Radiator Spring Racers, which has a wait time R uniformly distributed between 0.0 and 1.0.
- Sally chooses Space Mountain, which has a wait time S that follows an exponential PDF with parameter $\lambda = 2$.
- Tom chooses Toy Story Mania, which has a wait time T that follows the distribution defined by the following PDF:

$$f_T(t) = \begin{cases} 2-2t & \text{if } t \in [0,1] \\ 0 & \text{elsewhere.} \end{cases}$$

Note that R, S, and T are all *continuous* and mutually independent random variables that have units of hours.

(a) (6 pts) **Determine $\mathbb{E}[T]$ and give an explicit expression for the CDF $F_T(t) = \Pr(T \leq t)$. Show all your work for full credit.** Be sure that your answer for $F_T(t)$ is well-defined for all real number inputs.

(b) (5 pts) Let A be the event that the wait time for Space Mountain is greater than h hours. **Determine $f_{S|A}(s)$, the conditional PDF of S given A, in terms of s and h. Show all your work for full credit.** Be sure that your answer is well-defined for all real number inputs s and h.

Also answer the following: **If Sally has already waited h hours, how much more time is she expected to wait before getting on the ride?** Justify your answer.

(c) (8 pts) Let W be the random variable which represents the wait time (in hours) of the person (out of Rohan, Sally, and Tom) who got onto their respective ride *last*. **Explicitly determine $F_W(w)$, the CDF of W. Show all your work for full credit.** Be sure that your answer is well-defined for all real number inputs w.

(HINT: In addition to what you found in part (a), find $F_R(r)$ and $F_S(s)$, the CDFs of R and S, respectively. Doing so will earn partial credit as well as guide you towards the answer.)

© UCB EECS 70, Spring 2020. All Rights Reserved. This may not be publicly shared without explicit permission.

4. Transparent Coupon Collecting (10 pts)

Lalitha decides to go collecting coupons from transparent cereal boxes. There are 10 unique kinds of coupons and Lalitha wishes to collect one of each kind. Each shop has 5 cereal boxes in stock, and each box has exactly one coupon in it. Each box's coupon is independently randomly chosen, with the 10 kinds of coupons being equally likely. Lalitha can see which coupons are in which boxes before picking which one box to buy. Due to the current restrictions in place, she may buy at most one cereal box per shop and can't visit the same shop more than once. Still, Lalitha must collect all kinds of coupons, and wishes to *minimize* the number of shops visited.

What is the expected number of shops Lalitha visits? Show all your work for full credit. You may leave your answer as a summation.

(HINT: If Lalitha already has k distinct coupon types before walking into a shop, what is the probability that Lalitha does not find a new coupon type in the shop?)

© UCB EECS 70, Spring 2020. All Rights Reserved. This may not be publicly shared without explicit permission.

5. Trios (22 pts)

Suppose there are n students at a Homework Party. A TA walks around and gives each student a wristband that is either red, blue, or green, with equal probability for each. The students' wristband colors are independent from one another.

For this question, a "trio" is defined as an unordered *set* of three students. (For instance, when $n = 4$, if we think of the students as being A, B, C, D; then there are four different trios: ABC, ABD, ACD, BCD.)

A "same-color trio" is a trio in which all three members of the trio receive the same wristband color.

Let S be the random variable representing the number of same-color trios among n students, where $n \geq 6$ throughout this problem.

(a) (5 pts) **What is $\mathbb{E}[S]$ in terms of n? Show all your work for full credit.**
(HINT: Write S in terms of a sum of indicator random variables. What does each indicator random variable represent? How many possible trios of students are there, in terms of n? What is the probability that the specific trio ABC is a same-color trio? Answering these will earn partial credit as well as guide you towards the answer.)

(b) (5 pts) With respect to any given trio (e.g. ABC), we can classify any other trio into one of three groups, based on the number of members it shares with the given trio. Let t_k be the number of trios that share k members with the given trio, where $k \in \{0,1,2\}$. (No need to define $t_3 = 1$ because there is only one trio that shares 3 members with a given trio, since three members define a unique trio.) Clearly, if t is the total number possible trios, $t = t_0 + t_1 + t_2 + 1$.

In terms of the total number of people n, determine an expression for each of t_0, t_1, and t_2. Justify your answer for full credit.

(c) (2 pts) Assume we know that the trio ABC is a same-color trio. **Given that, what is the probability that the trio ABD is also a same-color trio?** Justify your answer for full credit.

(d) (10 pts) **What is $\mathbb{E}[S^2]$ in terms of n? Show all your work for full credit.** Whenever possible, use the terms t_0, t_1, and t_2 from part (b) instead of plugging in the actual expressions.
(HINT: You're going to have to expand out S^2 and then group terms by category. After you do that, the previous two parts can be helpful.)
(HINT 2: This part can take some time to work out. Unless you're confident, we suggest that you make sure you've attempted all problems before doing this part.)

© UCB EECS 70, Spring 2020. All Rights Reserved. This may not be publicly shared without explicit permission.

6. Meow at the Market (18 pts)

Wendy and Earnest lost their cat Meow at the flea market. They remember Meow's last position, which was at the candy stall, and know that in the i-th minute after Meow went missing, Meow moved east by X_i meters, where $X_i \sim \text{Poisson}(2)$ and all X_i are independent and identically distributed (i.i.d.). It then makes sense to consider $S_n = \sum_{i=1}^{n} X_i$, which can be interpreted as the distance Meow is from the stand after n minutes.

Recall that for a Poisson(λ) random variable, the mean is λ and the variance is also λ.

(a) (4 pts) **Determine** $\mathbb{E}[S_n]$**, and Var**(S_n)**. Justify your answers for full credit.**

For each of the remaining parts, assume that $n = 100$ (i.e. 100 minutes have elapsed).

To find Meow, Wendy and Earnest construct a search interval which is centered around the position located $\mathbb{E}[S_{100}]$ meters east of the candy stall.

(b) (6 pts) Using Chebyshev's inequality, **provide the tightest upper bound on the width** W **of the interval such that we can guarantee Meow is within the search interval with probability at least** $\frac{7}{8}$**. Show all your work for full credit.**

(c) (8 pts) Say Wendy and Earnest want to change the width of their interval but decide to keep it centered at the same location: $\mathbb{E}[S_{100}]$ meters east of the candy stall. They want their chances of finding Meow in the search interval to be approximately 97%. Using the Central Limit Theorem and the standard normal CDF table (on the next page) as a means of approximation, **what width** W **should they choose for their search interval? Show all your work for full credit.**

© UCB EECS 70, Spring 2020. All Rights Reserved. This may not be publicly shared without explicit permission.

Introduction to Probability, 2nd Ed, by D. Bertsekas and J. Tsitsiklis, Athena Scientific, 2008

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

The standard normal table. The entries in this table provide the numerical values of $\Phi(y) = \mathbf{P}(Y \leq y)$, where Y is a standard normal random variable, for y between 0 and 3.49. For example, to find $\Phi(1.71)$, we look at the row corresponding to 1.7 and the column corresponding to 0.01, so that $\Phi(1.71) = .9564$. When y is negative, the value of $\Phi(y)$ can be found using the formula $\Phi(y) = 1 - \Phi(-y)$.

© UCB EECS 70, Spring 2020. All Rights Reserved. This may not be publicly shared without explicit permission.