Math Review Notes—Probability

Gregory Faletto

Contents

1 Probability **4**

1.1 To Know for Math 505A Midterm 1 (Discrete Random Variables) 4

1.1.1 Definitions 4

1.1.2 Conditioning 5

1.1.3 Odds and Ends 6

1.1.4 Methods for Calculating Quantities 10

1.1.5 Discrete Random Variable Distributions 12

1.1.6 Indicator Method 16

1.1.7 Linear transformations of random variables 16

1.1.8 Poisson Paradigm (Poisson approximation for indicator method) 16

1.1.9 Asymptotic Distributions 16

1.2 Worked problems 17

1.2.1 Example Problems That Will Likely Appear on Midterm (and Final) 17

1.2.2 Problems we did in class that professor mentioned 21

1.2.3 Problems we did on homework 26

1.3 To Know for Math 505A Midterm 2 45

1.3.1 Definitions 45

1.3.2 Probability-Generating Functions 48

1.3.3 Moment-Generating Functions 48

1.3.4 Characteristic Functions 48

1.3.5 Continuous Random Variable Distributions 49

1.3.6 Multivariate Gaussian (Normal) Distributions 54

1.4 Worked problems 55

1.4.1 Example Problems That Will Likely Appear on Midterm (and Final) 55

1.4.2 More Problems From Homework 61

Last updated January 19, 2019

1 Probability

These are my notes from taking Math 505A at USC, Math 541A at USC, ISE 620 at USC and the textbook *Probability and Random Processes* (Grimmett and Stirzaker) 3rd edition.

1.1 To Know for Math 505A Midterm 1 (Discrete Random Variables)

1.1.1 Definitions

Definition 1.1. The **probability mass function** of a discrete random variable X is the function $f : \mathbb{R} \to [0,1]$ given by $f(x) = \Pr(X = x)$.

Definition 1.2. The **(cumulative) distribution function** of a discrete random variable F is given by

$$F(x) = \sum_{i: x_i \leq x} f(x_i)$$

Definition 1.3. The **joint probability mass function** $f : \mathbb{R}^2 \to [0,1]$ of two discrete random variables X and Y is given by

$$f(x, y) = \Pr(X = x \cap Y = y)$$

Definition 1.4. The **joint distribution function** $F : \mathbb{R}^2 \to [0,1]$ is given by

$$F(x, y) = \Pr(X \leq x \cap Y \leq y)$$

Definition 1.5. If $\Pr(B) > 0$ then the **conditional probability** that A occurs given that B occurs is defined to be

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Definition 1.6. (Independent sets.) Let $A_1, A_2, \ldots$ be subsets of a sample space Ω, and let $\mathbb{P}$ be a probability law on Ω. We say that $A_1, A_2, \ldots$ are **independent** if for any finite subset S of $\{1, 2, \ldots\}$, we have

$$\mathbb{P}\left(\bigcap_{i \in S} A_i \right) = \prod_{i \in S} \mathbb{P}(A_i)$$

Definition 1.7. (Independence of random variables.) Random variables $X_1, X_2, \ldots$ are **independent** if for every $B_1, B_2, \ldots \subseteq \mathbb{R}$, the events $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \ldots$ are independent; that is,

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in B_i\} \right) = \prod_{i=1}^{n} \mathbb{P}(\{X_i \in B_i\})$$

Remark. A more informal definition is as follows: Two random variables X and Y are **independent** if and only if $\Pr(X \cap Y) = \Pr(X) \Pr(Y)$.

Theorem 1. (Law of total probability). If X is a random variable and Y is a discrete random variable taking on values $y_1, y_2, \ldots, y_n$, then $\Pr(X) = \sum_i \Pr(X \mid Y = y_i) \cdot \Pr(Y = y_i)$. (Can be used to prove independence.)

Definition 1.8. Two random variables X and Y are **uncorrelated** if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Proposition 2. (a) Two random variables are uncorrelated if and only if their covariance $\text{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ equals 0.

(b) If X and Y are independent then they are uncorrelated.

Theorem 3. If X and Y are independent and $g, h : \mathbb{R} \to \mathbb{R}$, then $g(X)$ and $h(Y)$ are also independent.

1.1.2 Conditioning

Definition 1.9. The **conditional distribution function** of Y given $X = x$, written $F_{Y|X}(\cdot \mid x)$, is defined by

$$F_{Y|X}(y \mid x) = \Pr(Y \leq y \mid X = x)$$

Definition 1.10. The **conditional probability mass function** of Y given $X = x$, written $f_{Y|X}(\cdot \mid x)$, is defined by

$$f_{Y|X}(y \mid x) = \Pr(Y = y \mid X = x)$$

Theorem 4. Iterated expectations:

(i) $\mathbb{E}\big[\mathbb{E}(X \mid Y)\big] = \mathbb{E}(X)$ **(Law of Total Expectation)**

(ii) $\mathbb{E}\big[(X \mid Y) \mid Z\big] = \mathbb{E}(X \mid Y)$

(iii) $\mathbb{E}(E(XY \mid Y)) = \mathbb{E}(Y\mathbb{E}(X \mid Y))$

Proof. (i) Discrete case:

$$\mathbb{E}\big[\mathbb{E}(X \mid Y)\big] = \sum_y \mathbb{E}(X \mid Y = y) \Pr(Y = y) = \sum_y \sum_x x \Pr(X = x \mid Y = y) \Pr(Y = y)$$

$$= \sum_y \sum_x x \Pr(X = x \cap Y = y) = \sum_x x \sum_y \Pr(X = x \cap Y = y) = \sum_x x \Pr(X = x) = \mathbb{E}(X)$$

Continuous case:

$$\mathbb{E}\big[\mathbb{E}(X \mid Y)\big] = \int_{-\infty}^{\infty} \mathbb{E}(X \mid Y = y) f_Y(y) dy = \text{ (by definition 1.75) } \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx \right) f_Y(y) dy$$

$$\text{(by Fubini's Theorem, Theorem \textbf{??})} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) f_Y(y) dy \right) dx = \int_{-\infty}^{\infty} x f_X(x) dx = \mathbb{E}(X)$$

□

Definition 1.11. Conditional Variance: $\text{Var}(X \mid Y) = \mathbb{E}\big[(X - \mathbb{E}(X \mid Y))^2 \mid Y\big]$

1.1.3 Odds and Ends

Proposition 5. Inclusion-Exclusion Principle:

(a)

$$\Pr\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \left(\sum_{1 \leq i_1 < \ldots < i_k \leq m} \Pr\left(A_{i1} \cap \ldots \cap A_{ik}\right)\right)$$

(b)

$$\left|\bigcup_{i=1}^n A_i\right| = \sum_{k=1}^n (-1)^{k+1} \left(\sum_{1 \leq i_1 < \ldots < i_k \leq m} |A_{i1} \cap \ldots \cap A_{ik}|\right)$$

To prove this, we will first prove the Multi-Binomial Theorem.

Lemma 6. (Multi-Binomial Theorem)

$$\prod_{i=1}^d (x_i + y_i)^{n_i} = \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} \cdots \sum_{k_d=0}^{n_d} \binom{n_1}{k_1} x_1^{k_1} y_1^{n_1 - k_1} \binom{n_2}{k_2} x_2^{k_2} y_2^{n_2 - k_2} \cdots \binom{n_d}{k_d} x_d^{k_d} y_d^{n_d - k_d}$$

Proof. □

We are now ready to prove the Inclusion-Exclusion Principle.

Proof. (Proof of (a).) Begin by noting that

$$\mathbf{1}_{\{\cup_{i=1}^n A_i\}} = 1 - \prod_{i=1}^n (1 - \mathbf{1}_{\{A_i\}}) \tag{1}$$

because the expression on the right will equal 1 if at least one term in the product equals 0 (that is, if $\mathbf{1}_{\{A_i\}} = 1$ for some $i \in 1, \ldots, n$) and will equal 0 if every term in the product equals 1 (if $\mathbf{1}_{\{A_i\}} = 0$ for every $i \in 1, \ldots, n$), which is exactly what we want. Expanding the right side of (1) using the Multi-Binomial Theorem (Lemma 6), we have

$$= 1 - \prod_{i=1}^n (1 - \mathbf{1}_{\{A_i\}}) = 1 - (1 - \mathbf{1}_{\{A_1\}})(1 - \mathbf{1}_{\{A_2\}}) \cdots (1 - \mathbf{1}_{\{A_n\}})$$

$$= 1 - \left[1 + \sum_{k=1}^{n} (-1)^k \left(\sum_{1 \leq i_1 < \ldots < i_k \leq k} \mathbf{1}_{\{A_{i_1}\}} \cdots \mathbf{1}_{\{A_{i_k}\}} \right)\right] = -1 \cdot \sum_{k=1}^{n} (-1)^k \left(\sum_{1 \leq i_1 < \ldots < i_k \leq k} \mathbf{1}_{\{A_{i_1} \cap \ldots \cap A_{i_k}\}} \right)$$

$$= \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < \ldots < i_k \leq k} \mathbf{1}_{\{A_{i_1} \cap \ldots \cap A_{i_k}\}} \right)$$

$$\implies \mathbf{1}_{\{\cup_{i=1}^n A_i\}} = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < \ldots < i_k \leq k} \mathbf{1}_{\{A_{i_1} \cap \ldots \cap A_{i_k}\}} \right) \tag{2}$$

Taking expectations of both sides of (2) yields

$$\Pr(\cup_{i=1}^n A_i) = \mathbb{E}\left[\sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < \ldots < i_k \leq k} \mathbf{1}_{\{A_{i_1} \cap \ldots \cap A_{i_k}\}} \right)\right]$$

$$= \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < \ldots < i_k \leq k} \mathbb{E}[\mathbf{1}_{\{A_{i_1} \cap \ldots \cap A_{i_k}\}}] \right)$$

$$= \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < \ldots < i_k \leq k} \Pr\left(A_{i_1} \cap \ldots \cap A_{i_k}\right) \right)$$

$\square$

Theorem 7. Sums of random variables. If X and Y are independent then

$$\Pr(X + Y = z) = f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x) = \sum_y f_X(z - y) f_Y(y)$$

Proposition 8. Variance-Covariance Expansion. Let $X_1, \ldots, X_n$ be random variables. If $\mathbb{E}\,|X_k|^2 < \infty$, then

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_k \mathrm{Var}(X_k) + \sum_{k \neq m} \sum_m \mathrm{Cov}(X_k, X_m)$$

Definition 1.12. Let $\{X_i\}$ be i.i.d. random variables. Let N be a random variable taking on positive integer values. Let

$$S = \sum_{i=1}^{N} X_i$$

Then S is a **compound random variable**.

Proposition 9. (Wald's Equation.) Let $\{X_i\}$ be i.i.d. random variables with mean $\mathbb{E}(X)$. Let N be a random variable taking on positive integer values, and let $S = \sum_{i=1}^{N} X_i$. Then $\mathbb{E}(S) = \mathbb{E}(N)\mathbb{E}(X)$.

Proof.

$$\mathbb{E}(S \mid N) = \mathbb{E}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \mathbb{E}(X_i) = N\mathbb{E}(X)$$

$$\mathbb{E}(S) = \mathbb{E}(\mathbb{E}[S \mid N]) = \mathbb{E}(N\mathbb{E}(X)) = \mathbb{E}(N)\mathbb{E}(X)$$

□

Proposition 10. (Proposition 1.6.1 in Sheldon Ross *A First Course in Probability.*) There are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, x_2, \ldots, x_r), x_i > 0 \ \forall \ i$ satisfying the equation $x_1 + x_2 + \ldots + x_r = n$.

Proof. (Not rigorous, but a justification.) Imagine we have n indistinguishable objects to allocate to r people. We lay out the n objects and take $r-1$ sticks to place in the $n-1$ spaces between them. The first person gets all the objects to the left of the leftmost stick, the second person gets the objects between the leftmost and second leftmost stick, and so on, until the last person gets all the objects to the right of the rightmost stick. The constraint that x_i be positive is equivalent to saying that each person must receive at least one object. Therefore we must place each stick in a different place. There are $\binom{n-1}{r-1}$ ways to do this.

□

Proposition 11. (Proposition 1.6.2 in Sheldon Ross *A First Course in Probability.*) There are $\binom{n+r-1}{r-1}$ distinct nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r), x_i \geq 0 \ \forall \ i$ satisfying the equation $x_1 + x_2 + \ldots + x_r = n$.

Proof. We would like to solve the problem

$$x_1 + x_2 + \ldots + x_r = n, x_i \geq 0 \ \forall i$$

Note that we can transform this problem in the following way:

$$x_1 + 1 + x_2 + 1 + \ldots + x_r + 1 = n + 1 \cdot r, x_i + 1 \geq 1 \ \forall i$$

Letting $y_i = x_i + 1$, we have the equivalent system

$$y_1 + y_2 + \ldots + y_r = n + r, y_i \geq 1 \ \forall i$$

Since $y_i \geq 1 \iff y_i > 0$, by Proposition 10, the number of distinct solutions to this equation is $\binom{n+r-1}{r-1}$.

□

Proposition 12. More generally, if we desire solutions to $x_1 + x_2 + \ldots + x_r = n$ such that $x_i \geq k \in \mathbb{N}$, then $\binom{n+r\cdot(1-k)-1}{r-1} = \binom{n+r-1-rk}{r-1}$ solutions are possible.

Proof. We can construct a similar argument to that used in the proof of Proposition 11 by adding $r\cdot(1-k)$ to each side of $x_1+x_2+\ldots+x_r=n$:

$$x_1+1-k+x_2+1-k+\ldots+x_r+1-k=n+r\cdot(1-k),\ x_i\geq k\ \forall i$$

Then substitute $y_i=x_i+1-k$ to yield

$$y_1+y_2+\ldots+y_r=n+r\cdot(1-k),\ y_i+k-1\geq k\ \forall i$$

then apply Proposition 10 noting that $y_i+k-1\geq k \iff y_i\geq 1 \iff x_i\geq k$ to yield the result. $\square$

Proposition 13. Even more generally, suppose we desire solutions to $x_1+x_2+\ldots+x_r=n$ such that $x_1\geq k_1\in\mathbb{N}, x_2\geq k_2\in\mathbb{N},\ldots,x_r\geq k_r\in\mathbb{N}$. Then

$$\binom{n+\sum_{i=1}^{r}(1-k_i)-1}{r-1}=\binom{n+r-1-\sum_{i=1}^{r}k_i}{r-1}$$

solutions are possible.

Proof. Very similar to the proof of Proposition 12. Add $\sum_{i=1}^{r}(1-k_i)$ to each side of $x_1+x_2+\ldots+x_r=n$:

$$x_1+1-k_1+x_2+1-k_2+\ldots+x_r+1-k_r=n+\sum_{i=1}^{r}(1-k_i),\ x_i\geq k_i\ \forall i$$

Then substitute $y_i=x_i+1-k_i$ to yield

$$y_1+y_2+\ldots+y_r=n+\sum_{i=1}^{r}(1-k_i),\ y_i+k_i-1\geq k_i\ \forall i$$

Finally, apply Proposition 10 noting that $y_i+k_i-1\geq k_i \iff y_i\geq 1 \iff x_i\geq k_i$ to yield the result. $\square$

Proposition 14. Suppose we desire solutions to $x_1+x_2+\ldots+x_r=n$ such that $\tau\leq r$ of the $\{x_i\}$ exceed some threshold k. For example, if we have $x_1\geq k, x_2\geq k,\ldots,x_\tau\geq k$, with $x_{\tau+1},\ldots,x_r$ taking on arbitrary values, then the condition is satisfied. (The particular x_i that exceed k does not matter, as long as τ of them exceed k). Then

$$\binom{r}{\tau}\binom{n+r-1-k\tau}{r-1}$$

solutions are possible.

Proof. By Proposition 13, the number of ways this condition can be met for a particular set of τ variables x_i is $\binom{n+r-1-k\tau}{r-1}$. Since there are $\binom{r}{\tau}$ ways to choose which τ variables will exceed k, the result follows. $\square$

Remark. Convolution on the integers. Let X, Y be independent integer-valued random variables. Let $t \in \mathbb{Z}$.

$$\Pr(X+Y=t) = \sum_{j,k \in \mathbb{Z}: j+k=t} \Pr(X=j, Y=k) = \sum_{j \in F} \Pr(X=j, Y=t-j) = \sum_{j \in \mathbb{Z}} \Pr(X=j) \Pr(Y=t-j)$$

$$= \sum_{j \in \mathbb{Z}} p_X(j) p_Y(t-j)$$

Definition 1.13. Let $g, h : \mathbb{Z} \to \mathbb{R}$ be functions. The **convolution** of g and h, denoted by $g * h$, is a function $g * h : \mathbb{Z} \to \mathbb{R}$ defined by

$$(g * h)(t) = \sum_{j \in \mathbb{Z}} g(j) h(t-j) \quad \forall t \in \mathbb{Z}$$

Definition 1.14. (Convolution on the real line.) Let $g, h : \mathbb{R} \to \mathbb{R}$ b functions. The **convolution** of g and h, denoted by $g * h$, is a function $g * h : \mathbb{R} \to \mathbb{R}$ defined by

$$(g * h)(t) = \int_{-\infty}^{\infty} g(x) h(t-x) dx, \quad \forall t \in \mathbb{R}$$

Proposition 15. Let X, Y be two continuous independent random variables such that $\Pr(X+Y \leq t)$ is differentiable with respect to $t \in \mathbb{R}$. Then

$$f_{X+Y}(t) = (f_X * f_Y)(t), \quad \forall t \in \mathbb{R}$$

Proof.

$$\Pr(X+Y \leq t) = \int_{\{(x,y) \in \mathbb{R}^2 : x+y \leq t\}} f_{X,Y}(x,y) dx dy = \int_{x=-\infty}^{x=\infty} \int_{y=-\infty}^{y=t-x} f_X(x) f_Y(y) dy dx$$

Then, since $\Pr(X+Y \leq t)$ is differentiable with respect to t, we have by the Fundamental Theorem of Calculus

$$f_{X+Y}(t) = \frac{\mathrm{d}}{\mathrm{d}t} \Pr(X+Y \leq t) = \int_{x=-\infty}^{x=\infty} f_X(x) \frac{\mathrm{d}}{\mathrm{d}t} \int_{y=-\infty}^{y=t-x} f_Y(y) dy dx = \int_{x=-\infty}^{x=\infty} f_X(x) f_Y(t-x) dx$$

□

1.1.4 Methods for Calculating Quantities

- Expectation
  - **Definition 1.15.** $\mathbb{E}(X) = \sum_x x \Pr(X=x)$
  -

Theorem 16. (a) $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
(b) If $X \geq 0$ then $\mathbb{E}(X) \geq 0$

–

Theorem 17. Law of the Unconscious Statistician: If X has mass function f, and $g : \mathbb{R} \to \mathbb{R}$, then

$$\mathbb{E}(g(X)) = \sum_x g(x)f(x)$$

–

Proposition 18. Expectation is a linear operator: $\mathbb{E}\big(\sum_i X_i\big) = \sum_i \mathbb{E}(X_i)$

- Variance

–

Definition 1.16. $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$

–

Proposition 19. (Useful reformulation:) $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

–

Theorem 20. (Some useful results):

(a) $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
(b) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$
(c) $\mathrm{Var}(aX \pm bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \pm 2ab\mathrm{Cov}(X, Y)$
(d) $\mathrm{Var}\Bigg(\sum_{i=1}^n X_i\Bigg) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i, X_j)$

–

Definition 1.17. Conditional variance:

$$\mathrm{Var}(X \mid Y) = \mathbb{E}\big[(X - \mathbb{E}(X \mid Y))^2 \mid Y\big]$$

–

Theorem 21. Law of Total Variance: $\mathrm{Var}(X) = \mathrm{Var}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}\big(\mathrm{Var}(X \mid Y)\big)$

Proof.

$$\mathrm{Var}\big(\mathbb{E}(X \mid Y)\big) = \mathbb{E}[(\mathbb{E}(X \mid Y))^2] - [\mathbb{E}(\mathbb{E}(X \mid Y))]^2 = \mathbb{E}[(\mathbb{E}(X \mid Y))^2] - \mathbb{E}(X)^2 \tag{3}$$

$$\mathrm{Var}(X \mid Y) = \mathbb{E}(X^2 \mid Y) - (\mathbb{E}(X \mid Y))^2 \implies \mathbb{E}[\mathrm{Var}(X \mid Y)] = \mathbb{E}(X^2) - \mathbb{E}[(\mathbb{E}(X \mid Y))^2] \tag{4}$$

Adding together (3) and (4) yields

$$\mathrm{Var}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}\big(\mathrm{Var}(X \mid Y)\big) = \mathbb{E}[(\mathbb{E}(X \mid Y))^2] - \mathbb{E}(X)^2 + \mathbb{E}(X^2) - \mathbb{E}[(\mathbb{E}(X \mid Y))^2]$$

$$= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathrm{Var}(X)$$

□

Corollary 21.1. (Rao-Blackwell Theorem.) $\text{Var}(X) \geq \text{Var}(\mathbb{E}(X \mid Y))$

Proof. Follows immediately from Theorem 21 by noting that since the variance is nonnegative, $\mathbb{E}\big(\text{Var}(X \mid Y)\big) \geq 0$.

□

–

Proposition 22. If $c \in \mathbb{R}$, then $\text{Var}(c) = 0$.

- Covariance

  –

  Definition 1.18. $\text{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big]$

  –

  Proposition 23. (Useful reformulation): $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

  –

  Theorem 24. (Some useful results):

  (a) $\text{Cov}(aX, BY) = ab\text{Cov}(X, Y)$
  (b) $\text{Cov}(X, X) = \text{Var}(X)$
  (c) $\text{Cov}(aX + bY) = ac\text{Var}(X) + bd\text{Var}(Y) + (ad + bc)\text{Cov}(X, Y)$
  (d) $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$
  (e) $\text{Cov}\left(\sum_{i=1}^n X_i, Y\right) = \sum_{i=1}^n \text{Cov}(X_i, Y)$
  (f) $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$

  –

  Definition 1.19. Conditional covariance:

  $$\text{Cov}(X, Y \mid Z) = \mathbb{E}(XY \mid Z) - \mathbb{E}(X \mid Z)\mathbb{E}(Y \mid Z) = \mathbb{E}\big[(X - \mathbb{E}(X \mid Z))(Y - \mathbb{E}(Y \mid Z)) \mid Z\big]$$

  –

  Theorem 25. Law of Total Covariance:

  $$\text{Cov}(X, Y) = \mathbb{E}\big(\text{Cov}(X, Y \mid Z)\big) + \text{Cov}\big(\mathbb{E}(X \mid Z), \mathbb{E}(Y \mid Z)\big)$$

1.1.5 Discrete Random Variable Distributions

Binomial: $\text{Binomial}(n, p)$ (sum of n Bernoulli random variables)

- Mass function: $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Distribution: $\Pr(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}$
- Expectation: $\mathbb{E}(X) = np$
- Variance: $\text{Var}(X) = np(1-p)$

Poisson: Poisson(λ): an approximation of the binomial distribution for n very large, p very small, $np \to \lambda \in (0, \infty)$.

- Mass function:
$$\Pr(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$
- Distribution: $\Pr(X \leq k) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^i}{i!}$
- Expectation: $\mathbb{E}(X) = \lambda$ (derive from basic definitions)
- Variance: $\text{Var}(X) = \lambda$

Proposition 26. Let $X \sim \text{Binomial}(n, p)$. Then

$$\lim_{n\to\infty,\ p\to 0,\ np\to\lambda} X \sim \text{Poisson}(np).$$

Proof.
$$\lim_{n\to\infty,\ p\to 0,\ np\to\lambda} \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n\to\infty,\ p\to 0,\ np\to\lambda} \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

$$= \frac{1}{k!} \cdot \lim_{n\to\infty,\ p\to 0,\ np\to\lambda} (1-p)^{n-k} p^k \prod_{i=0}^{k-1} (n-i) = \frac{1}{k!} \cdot \lim_{n\to\infty,\ p\to 0,\ np\to\lambda} \left(1 - \frac{np}{n}\right)^{n-k} p^k \prod_{i=0}^{k-1} (n-i)$$

Using $\lim_{n\to\infty}(1 - \lambda/n)^n = \exp(\lambda)$, and letting $\lambda = np$, we have

$$= \frac{\exp(-np)(np)^k}{k!} = \boxed{\frac{\exp(-\lambda)\lambda^k}{k!}} \sim \text{Poisson}(np)$$

□

Proposition 27. Let $X \sim \text{Poisson}(\lambda)$. If λ is sufficiently large (say $\lambda > 20$, then we can use the approximation

$$X \sim \mathcal{N}(\lambda, \lambda)$$

Proof. (Informal justification.) By Proposition 26, a Poisson distribution can be thought of as a close approximation of a binomial distribution. Since a binomial distribution can be approximated by a normal distribution for n large and np not too large, the same is true of a Poisson distribution. □

Geometric: $\text{G}_1(p)$: the number of Bernoulli trials before the first success.

- Mass function: $\Pr(X = k) = p(1-p)^{k-1}$
- Distribution: $\Pr(X \leq k) = \sum_{i=1}^{k} p(1-p)^{k-1}$
- Expectation: $\mathbb{E}(X) = 1/p$

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

Negative binomial: $\mathrm{NB}(r, p)$: The number of Bernoulli trials required for r successes. (Can be derived as the sum of r identically distributed geometric random variables.)

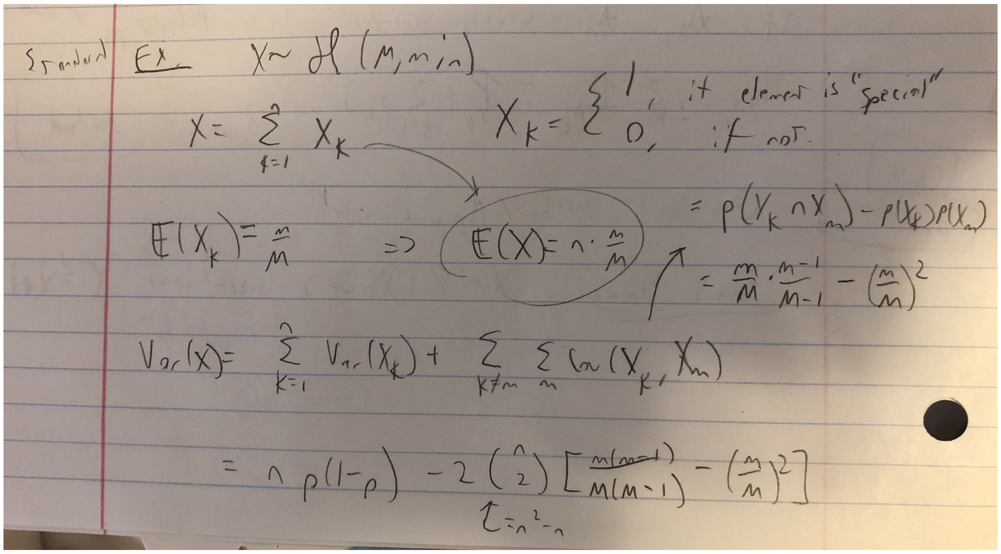
- Mass function: $\Pr(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$
- Distribution: $\Pr(X \leq k) = \sum_{i=r}^{k} \binom{i-1}{r-1} p^r (1-p)^{i-r}$
- Expectation: $\mathbb{E}(X) = \frac{pr}{1-p}$
- Variance: $\mathrm{Var}(X) = \frac{pr}{(1-p)^2}$

Hypergeometric: Hypergeometric(N, M, K): When drawing a sample of size K from a group of N items, M of which are special, X is the number of special items retrieved.

- Mass function:
$$\Pr(X = k) = \frac{\binom{M}{k}\binom{N-M}{K-k}}{\binom{N}{K}}$$
- Distribution:
$$\Pr(X \leq k) = \sum_{i=0}^{k} \frac{\binom{M}{i}\binom{N-M}{K-i}}{\binom{N}{K}}$$
- Expectation: $\mathbb{E}(X) = \frac{MK}{N}$

  Proof. Let Y_j be an indicator variable for special item j being selected. Note that $X = \sum_{j=1}^{M} Y_j$ and that
$$\mathbb{E}(Y_j) = \frac{\binom{1}{1}\binom{N-1}{K-1}}{\binom{N}{K}} = \frac{1 \cdot \frac{(N-1)!}{(N-K)!(K-1)!}}{\frac{N!}{(N-K)!K!}} = \frac{K}{N},$$
  so we have
$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{j=1}^{M} Y_j\right) = \sum_{j=1}^{M} \mathbb{E}(Y_j) = \frac{MK}{N}.$$
  Alternative proof:

  Let Z_i be an indicator variable for the ith selected item to be special. Then $X = \sum_{i=1}^{K} Z_i$ and $\mathbb{E}(Z_i) = \frac{M}{N}$, so we have
$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{K} Z_i\right) = \sum_{i=1}^{K} \mathbb{E}(Z_i) = \frac{MK}{N}.$$
  Variance: $\mathrm{Var}(X) =$ (find by indicator method. Proof from notes, I think this may have errors:)



Standard Ex. $X \sim \mathcal{H}(M, m, n)$

$X = \sum_{k=1}^{n} X_k$ $\quad X_k = \begin{cases} 1, & \text{if element is "special"} \\ 0, & \text{if not.} \end{cases}$

$\mathbb{E}(X_k) = \frac{m}{M} \Rightarrow \mathbb{E}(X) = n \cdot \frac{m}{M}$

$= P(X_k \cap X_m) - P(X_k)P(X_m)$

$= \frac{m}{M} \cdot \frac{m-1}{M-1} - \left(\frac{m}{M}\right)^2$

$\mathrm{Var}(X) = \sum_{k=1}^{n} \mathrm{Var}(X_k) + \sum_{k \neq m} \sum_{m} \mathrm{Cov}(X_k, X_m)$

$= n p(1-p) - 2 \binom{n}{2} \left[\frac{m(m-1)}{M(M-1)} - \left(\frac{m}{M}\right)^2\right]$

$= n^2 - n$

$\square$

Proposition 28. Let $X \sim \text{Hypergeometric}(N, M, K)$. Then

$$\lim_{M,N\to\infty, M/N\to p} \Pr(X = k) \sim \text{Binomial}(K, p)$$

Proof.

$$\lim_{M,N\to\infty,M/N\to p} p_k(M,N,K) = \lim_{M,N\to\infty,M/N\to p} \frac{\binom{M}{k}\binom{N-M}{K-k}}{\binom{N}{K}}$$

$$= \lim_{M,N\to\infty,M/N\to p} \frac{M!(N-M)!/[k!(M-k)!(K-k)!(N-M-K+k)!]}{N!/[K!(N-K)!]}$$

$$= \lim_{M,N\to\infty,M/N\to p} \frac{K!}{(K-k)!k!} \cdot \frac{M!(N-M)!(N-K)!}{N!(M-k)!(N-M-K+k)!}$$

$$= \lim_{M,N\to\infty,M/N\to p} \binom{K}{k} \cdot \frac{M!/(M-k)!}{N!/(N-k)!} \cdot \frac{(N-M)!(N-K)!}{(N-k)!(N-M-(K-k))!}$$

$$= \binom{K}{k} \lim_{M,N\to\infty,M/N\to p} \frac{M!/(M-k)!}{N!/(N-k)!} \cdot \frac{(N-M)!/(N-M-(K-k))!}{(N-K+(K-k))!/(N-K)!}$$

$$= \binom{K}{k} \lim_{M,N\to\infty,M/N\to p} \prod_{i=0}^{k-1} \frac{M-i}{N-i} \cdot \prod_{j=0}^{K-k-1} \frac{N-M-j}{N-K+1+j}$$

$$= \binom{K}{k} \left(\frac{M}{N}\right)^k \left(\frac{N-M}{N}\right)^{K-k}$$

$$= \binom{K}{k} \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{K-k} = \binom{K}{k} p^k (1-p)^{K-k}$$

$\square$

1.1.6 Indicator Method

Proposition 29. If $\mathbf{1}_{A_k}$ is an indicator then

(a)
$$\mathrm{Cov}(\mathbf{1}_{A_k}, \mathbf{1}_{A_m}) = \mathbb{E}(\mathbf{1}_{A_k}\mathbf{1}_{A_m}) - \mathbb{E}(\mathbf{1}_{A_k})\mathbb{E}(\mathbf{1}_{A_m}) = \Pr(A_k \cap A_m) - \Pr(A_k)\Pr(A_m)$$

(b)
$$\mathrm{Var}(\mathbf{1}_{A_k}) = \mathbb{E}(\mathbf{1}_{A_k}^2) = \mathbb{E}(\mathbf{1}_{A_k})^2 = \Pr(A_k) - (\Pr(A_k))^2$$

Theorem 30. X is independent of Y if and only if X is independent of $\mathbf{1}_A$, $A \in Y$.

Example problems: 505A Homework 3 problem 9(a)

Worked examples in p. 56 - 59 of Grimmett and Stirzaker 3rd edition.

1.1.7 Linear transformations of random variables

1.1.8 Poisson Paradigm (Poisson approximation for indicator method)

Theorem 31. (Theorem 4.12.9, p. 129 of Grimmett and Stirzaker.) Let A_i be an event. If $X = \sum_{i=1}^m \mathbf{1}_{A_i}$ where $\mathbf{1}_{A_i}$ is an indicator variable for A_i, and the A_i are only weakly dependent on each other, then

$$\text{As } m \to \infty, \quad X \sim \text{Poisson}(\mathbb{E}(X))$$

More specifically, let B_i be n independent Bernoulli random variables with probabilities p_i. If $Y = \sum_{i=1}^n B_i$ then

$$\text{As } n \to \infty, \quad Y \sim \text{Poisson}\left(\mathbb{E}\left(\sum_i B_i\right)\right) = \text{Poisson}\left(\sum_i \mathbb{E}B_i\right) = \text{Poisson}\left(\sum_i p_i\right)$$

Proof. Full proof available in Grimmett and Stirzaker, section 4.12, page 129. A justification of the first claim is as follows: if the A_i are independent and $\Pr(A_i) = p\ \forall i$, then $X \sim \text{Binomial}(m, p$. Then by Proposition 26, the result follows. It turns out that this result holds up if the probabilities are not necessarily identical (but all small) and the variables are not necessarily independent (but only weakly dependent). □

1.1.9 Asymptotic Distributions

Proposition 32.
$$e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n$$

Theorem 33. Stirling's Formula:

$$n! \approx n^n e^{-n}\sqrt{2\pi n}$$

1.2 Worked problems

1.2.1 Example Problems That Will Likely Appear on Midterm (and Final)

(1) (a) **Fall 2011 Problem 1 (same as HW1 problem 5; similar to HW3 problem 2(5).)** Let A and B be events such that $0 < \Pr(A) < 1$. Show that if $\Pr(B \mid A) = \Pr(B \mid A^c)$, then A and B are independent.

(b) Let X and Y be two discrete random variables, each taking only two possible values. Show that if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ then X and Y are independent.

Solution.

(a) A and B are independent if and only if

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

We know that

$$\Pr(B) = \Pr(B|A) \cdot \Pr(A) + \Pr(B|A^c) \cdot \Pr(A^c)$$

$$= \Pr(B|A) \cdot \Pr(A) + \Pr(B|A) \cdot (1 - \Pr(A)) = \Pr(B|A) \cdot \Pr(A) + \Pr(B|A) - \Pr(B|A) \cdot \Pr(A)$$

$$= \Pr(B|A)$$

Also, we know that since $\Pr(A) \neq 0$,

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

Per above $\Pr(B|A) = \Pr(B)$, so we have

$$\Pr(B) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

which is what we were trying to prove. So the answer is $\boxed{\text{true.}}$

(b) Without loss of generality, let X and Y have mass functions

$$X = \begin{cases} x_1 & \text{with probability } \Pr(A) \\ x_2 & \text{with probability } \Pr(A^c) \end{cases}$$

$$Y = \begin{cases} y_1 & \text{with probability } \Pr(B) \\ y_2 & \text{with probability } \Pr(B^c) \end{cases}$$

Then $X \perp\!\!\!\perp Y \iff \Pr(A \cap B) = \Pr(A)\Pr(B)$. Let $\alpha = X - x_2$, $\beta = Y - y_2$; that is,

$$\alpha = \begin{cases} x_1 - x_2 & \text{with probability } \Pr(A) \\ 0 & \text{with probability } \Pr(A^c) \end{cases}$$

$$\beta = \begin{cases} y_1 - y_2 & \text{with probability } \Pr(B) \\ 0 & \text{with probability } \Pr(B^c) \end{cases}$$

Then we have

- $\mathbb{E}(\alpha) = (x_1 - x_2)\Pr(A)$
- $\mathbb{E}(\beta) = (y_1 - y_2)\Pr(B)$
- $\mathbb{E}(\alpha\beta) = (x_1 - x_2)(y_1 - y_2)\Pr(A \cap B)$

which we can use to obtain

$$\mathbb{E}(XY) = \mathbb{E}\big[(\alpha + x_2)(\beta + y_2)\big] = \mathbb{E}(\alpha\beta) + y_2\mathbb{E}(\alpha) + x_2\mathbb{E}(\beta) + x_2 y_2$$

$$= (x_1 - x_2)(y_1 - y_2)\Pr(A \cap B) + y_2(x_1 - x_2)\Pr(A) + x_2(y_1 - y_2)\Pr(B) + x_2 y_2 \tag{5}$$

$$\mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(\alpha + x_2)\mathbb{E}(\beta + y_2) = \big[x_2 + (x_1 - x_2)\Pr(A)\big]\big[y_2 + (y_1 - y_2)\Pr(B)\big]$$

$$= x_2 y_2 + x_2(y_1 - y_2)\Pr(B) + y_2(x_1 - x_2)\Pr(A) + (x_1 - x_2)(y_1 - y_2)\Pr(A)\Pr(B) \tag{6}$$

Using $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, we set (5) and (6) equal to each other. Canceling terms appearing in both yields

$$(x_1 - x_2)(y_1 - y_2)\Pr(A \cap B) = (x_1 - x_2)(y_1 - y_2)\Pr(A)\Pr(B) \iff \Pr(A \cap B) = \Pr(A)\Pr(B)$$

which proves the independence of X and Y if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

(2) **Fall 2013 Qual Problem 1.** Consider a sequence of independent tosses of a pair of fair dice. Compute the probability that the sum 4 will occur before the sum 5.

Solution. Let Y_k be the outcome of the kth toss. Let X_{k1} be the number of the first die on the kth toss and X_{k2} be the outcome of the second die. Note that

$$\Pr(Y_k = 4) = \Pr(X_{k1} = 1 \cap X_{k2} = 3) + \Pr(X_{k1} = 2 \cap X_{k2} = 2) + \Pr(X_{k1} = 3 \cap X_{k2} = 1) = 3 \cdot \frac{1}{36} = \frac{1}{12}$$

$$\Pr(Y_k = 5) = \Pr(X_{k1} = 1 \cap X_{k2} = 4) + \Pr(X_{k1} = 2 \cap X_{k2} = 3) + \Pr(X_{k1} = 3 \cap X_{k2} = 2)$$

$$+ \Pr(X_{k1} = 4 \cap X_{k2} = 1) = 4 \cdot \frac{1}{36} = \frac{1}{9}$$

Let A_k be the event that $Y_k = 4$ and $Y_j \neq 4$ or $5,\ j = 1, \ldots, k-1$. Note that all A_k are mutually exclusive and

$$\Pr(A_k) = \frac{1}{12} \cdot \left(1 - \frac{3+4}{36}\right)^{k-1} = \frac{1}{12} \cdot \left(\frac{29}{36}\right)^{k-1}.$$

Then

$$\Pr(\{\text{roll a 4 before a 5}\}) = \Pr\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \Pr(A_k) = \frac{1}{12}\sum_{k=1}^{\infty}\left(\frac{29}{36}\right)^{k-1} = \frac{1/12}{1 - 29/36}$$

$$= \frac{3}{36 - 29} = \boxed{\frac{3}{7}}$$

(3) **Spring 2013 Problem 1, in notes from 09/21.** Let X and Y be random variables such that $\mathbb{E}(X \mid Y) = Y, \mathbb{E}(Y \mid X) = X, \mathbb{E}(X^2) < \infty, \mathbb{E}(Y^2) < \infty$. Show that $\mathbb{E}(X-Y)^2 = 0$ (or equivalently, show $\Pr(X = Y) = 1$).

Solution.

$$\mathbb{E}(X-Y)^2 = \mathbb{E}(X^2 - 2XY + Y^2) = \mathbb{E}(X^2) - 2\mathbb{E}(XY) + \mathbb{E}(Y^2)$$

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY \mid Y)) = \mathbb{E}(Y\mathbb{E}(X \mid Y)) = \mathbb{E}(Y \cdot Y) = \mathbb{E}(Y^2)$$

Also,

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY \mid X)) = \mathbb{E}(X\mathbb{E}(Y \mid X)) = \mathbb{E}(X \cdot X) = \mathbb{E}(X^2)$$

Therefore

$$\mathbb{E}(X-Y)^2 = 0$$

(4) **Fall 2016 Problem 4.** Consider a group of $n \geq 4$ people, among whom are Alice, Bob, Charles, and Diana, standing in a row. Assume that all possible orderings of the n people are equally likely.

(a) Compute the probability that Charles stands somewhere between Alice and Bob. (Note: this does not mean that the three are necessarily adjacent; there can be other people between Alice and Bob.)

(b) Compute the probability that Diana stands somewhere between Alice and Bob given that Charles stands somewhere between Alice and Bob.

(c) Let X be the number of people who stand between Alice and Bob. Compute the expected value and the variance of X. (Note: Alice and Bob themselves are not counted in this number.)

Solution.

(a) For this part, n does not make a difference; all we need to know is the ordering of A, B, and C. This is because conditional on a specific ordering of A, B, and C, all arrangements of everyone else are equally likely; and conversely, the a particular ordering of A, B, and C is independent of which three particular slots are made available to them. Examining the permutations of A, B, and C, two of them have C in the middle, so the answer is $2/6 = \boxed{1/3}$.

(b) Similarly, the answer is independent of n, so we work with $n = 4$. All possible orderings with Charles between Alice and Bob are as follows:

$$ACDB,\ ADCB,\ BCDA,\ BDCA,\ ACBD,\ BCAD,\ DACB,\ DBCA$$

The first four of these have Diana between Alice and Bob, so the answer is $4/8 = \boxed{1/2}$.

(c) Let I_k be an indicator variable for the event that person k is between A and B. By the result from part (a), $\mathbb{E}(I_k) = 1/3$. Then we have

$$\text{Var}(I_k) = \mathbb{E}(I_k^2) - \mathbb{E}(I_k)^2 = 1^2 \cdot \Pr(I_k = 1) - \frac{1}{9} = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$$

Noting that the four arrangements above with Charles and Diana in between Alice and Bob are the only ones where this will be the case of the $4! = 24$ possible orderings, we have

$$\mathbb{E}(I_k I_j) = \frac{4}{24} = \frac{1}{6}$$

so

$$\text{Cov}(I_j, I_k) = \mathbb{E}(I_k I_j) - \mathbb{E}(I_i)\mathbb{E}(I_j) = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$$

Therefore

$$\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{n-2} I_k) = \sum_{i=1}^{n-2} \frac{1}{3} = \frac{n-2}{3}$$

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n-2} I_k\right) = \sum_{i=1}^{n-2} \text{Var}(I_k) + 2 \sum_{1 \le j < k \le n-2} \text{Cov}(I_j, I_k) = \frac{2(n-2)}{9} + \left[(n-2)^2 - (n-2)\right] \cdot \frac{1}{18}$$

$$= \frac{4(n-2)}{18} + \frac{(n-2)(n-3)}{18} = \boxed{\frac{(n-2)(n+1)}{18}}$$

(5) **Spring 2015 Problem 2.** A deck of 52 cards is shuffled thoroughly. Someone goes through all 52 cards, scoring 1 point each time 2 cards of the same value are consecutive (that is, when two consecutive cards have the same rank but different suits). For example, the sequence 9H, 8H, 7D, 6C, 7S, 7H, 7C scores 2 points, one for the 7H following the 7S, one for the 7C following the 7H. Let X be the total score.

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

(b) Compute $\Pr(X = 39)$. (Note that there are 13 different ranks and you cannot score more than 3 per rank.)

(c) In the line below, circle the number that you think is the closest to the value $\Pr(X = 0)$ and briefly explain your choice:

$$\frac{1}{1000}, \quad \frac{1}{500}, \quad \frac{1}{100}, \quad \frac{1}{50}, \quad \frac{1}{20}, \quad \frac{1}{10}, \quad \frac{1}{5}, \quad \frac{1}{2}$$

Solution.

(a) Start by assuming the permutation is cyclic (that is, after the last card you go back to the beginning). Let Y be the number of matches in this situation. Let A_i be the event that the ith card is followed by a match. Then $\Pr(A_i) = 3/51 = 1/17$, so

$$Y = \sum_{k=1}^{52} \mathbf{1}_{\{A_i\}} \implies \mathbb{E}(Y) = \sum_{i=1}^{52} \mathbb{E}(\mathbf{1}_{\{A_i\}}) = \sum_{i=1}^{52} \Pr(A_i) = 52 \cdot 1/17 = 52/17$$

Note that $\mathbb{E}(Y) = \mathbb{E}(X) + \mathbb{E}(\mathbf{1}_{\{A_i\}})$ because if the permutation is cyclic then you have one extra opportunity to mach at the end.

$$\implies \mathbb{E}(X) = \frac{52}{17} - \frac{1}{17} = \frac{51}{17} = \boxed{3}$$

(b) $\Pr(X = 39) = \Pr(\{\text{all possible matches occur}\})$, so in this event all cards of the same rank are clustered together. There are 13 clusters of 4 cards, so there are 13! ways to order the clusters and 4! ways to order the cards within each cluster. Therefore

$$\boxed{\Pr(X = 39) = \frac{13!(4!)^{13}}{52!}}$$

(c) Because A_i are only weakly dependent and $\Pr(A_i)$ is small for all A_i, we can use the Poisson approximation (see Section 1.1.8); that is, $X \sim \text{Poisson}(\mathbb{E}(X)) = \text{Poisson}(3)$. Therefore

$$\Pr(X = 0) \approx \frac{e^{-3} \cdot 3^0}{0!} = \frac{1}{e^3} \approx \frac{1}{2.8^3} \approx \boxed{\frac{1}{20}}$$

1.2.2 Problems we did in class that professor mentioned

(Fall 2014 Problem 1) (Variation of Midterm problem 1 above) Let A and B be two events with $0 < \Pr(A) < 1,\ 0 < \Pr(B) < 1$. Define the random variables $\xi = \xi(\omega)$ and $\eta = \eta(\omega)$ by

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

True or false: the events A and B are independent if and only if the random variables ξ and η are uncorrelated?

Solution. ($\Longrightarrow$) Suppose A and B are independent. Then ξ and η are uncorrelated if and only if $\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta)$. We can write $\xi = 5 \cdot \mathbf{1}_A - 7 \cdot \mathbf{1}_{A^c}$ and $\eta = 2 \cdot \mathbf{1}_B + 3 \cdot \mathbf{1}_{B^c}$. So we have

$$\xi\eta = (5 \cdot \mathbf{1}_A - 7 \cdot \mathbf{1}_{A^c})(2 \cdot \mathbf{1}_B + 3 \cdot \mathbf{1}_{B^c}) = 10 \cdot \mathbf{1}_{A\cap B} + 15 \cdot \mathbf{1}_{A\cap B^c} - 14 \cdot \mathbf{1}_{A^c\cap B} - 21 \cdot \mathbf{1}_{A^c\cap B^c}$$

$$\implies \mathbb{E}(\xi\eta) = 10\Pr(A \cap B) + 15\Pr(A \cap B^c) - 14\Pr(A^c \cap B) - 21\Pr(A^c \cap B^c)$$

Then

$$\mathbb{E}(\xi)\mathbb{E}(\eta) = (5\Pr(A) - 7\Pr(A^c))(2\Pr(B) + 3\Pr(B^c))$$

$$= 10\Pr(A \cap B) + 15\Pr(A \cap B^c) - 14\Pr(A^c \cap B) - 21\Pr(A^c \cap B^c) = \mathbb{E}(\xi\eta)$$

where the second-to-last step follows from the independence of A and B. Therefore η and ξ are uncorrelated.

($\Longleftarrow$) Now suppose η and ξ are uncorrelated. Then ξ and η are independent if and only if $\Pr(\xi \cap \eta) = \Pr(\xi)\Pr(\eta)$. Define

$$\alpha(\omega) = \xi(\omega) + 7 = \begin{cases} 12 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}, \qquad \beta(\omega) = \eta(\omega) - 3 = \begin{cases} -1 & \text{if } \omega \in B \\ 0 & \text{if } \omega \notin B \end{cases}$$

Then we have

$$(\alpha\beta)(\omega) = \begin{cases} -12 & \text{if } \omega \in A \cap B \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mathbb{E}(\xi\eta) = \mathbb{E}\big[(\alpha - 7)(\beta + 3)\big] = \mathbb{E}(\alpha\beta) + 3\mathbb{E}(\alpha) - 7\mathbb{E}(\beta) - 21$$

$$\mathbb{E}(\xi)\mathbb{E}(\eta) = (\mathbb{E}(\alpha) - 7)(\mathbb{E}(\beta) + 3) = \mathbb{E}(\alpha)\mathbb{E}(\beta) - 7\mathbb{E}(\beta) + 3\mathbb{E}(\alpha) - 21$$

Since by assumption $\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta)$, this yields $\mathbb{E}(\alpha\beta) = \mathbb{E}(\alpha)\mathbb{E}(\beta)$. But

$$\mathbb{E}(\alpha\beta) = -12\Pr(A \cap B), \qquad \mathbb{E}(\alpha)\mathbb{E}(\beta) = 12\Pr(A)(-1)\Pr(B) = -12\Pr(A)\Pr(B)$$

Therefore $\Pr(\xi \cap \eta) = \Pr(\xi)\Pr(\eta)$ and ξ and η are independent.

Example (Letter/envelope matching problem; sometimes referred to as Montmort's matching problem). An assistant brings n sandwiches for n employees at a company. Each employee ordered a unique sandwich, but unfortunately the assistant forgot to ask that the sandwiches be labeled, so they are all indistinguishable, wrapped in the same paper. The assistant plans to distribute one sandwich to each employee and hope for the best. Let X be the number of sandwiches that are delivered to the correct person.

(a) What is the probability of at least one match; that is, $\Pr(X \geq 1)$?

(b) What is the probability of r correct matches?

(c) What $\mathbb{E}(X)$?

(d) What is $\mathrm{Var}(X)$?

Solution.

(a) Let A_k be an indicator variable for the event that sandwich k is matched to the correct employee. Then

$$\Pr(X \geq 1) = \Pr\left(\bigcup_{k=1}^{n} A_k\right)$$

Consider that if there are k correct matches, there are $\binom{n}{k}$ sets of k sandwiches that could be correctly distributed. Also, the probability of a particular set of k sandwiches being correctly distributed is $(n-k)!/n!$. So we have

$$\Pr(X = k) = \binom{n}{k} \frac{(n-k)!}{n!}$$

Therefore by the Inclusion-Exclusion Principle (Proposition 5),

$$\Pr\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} (-1)^{k-1} \Pr(X = k) = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k-1} \frac{n!}{(n-k)!k!} \frac{(n-k)!}{n!}$$

$$= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!} = \frac{(-1)^0}{0!} - \sum_{k=0}^{n} \frac{(-1)^k}{k!} = \boxed{1 - \sum_{k=0}^{n} \frac{(-1)^k}{k!}}$$

As $n \to \infty$, we have

$$1 - \sum_{k=0}^{n} \frac{(-1)^k}{k!} \to 1 - e^{-1} = \boxed{1 - \frac{1}{e}}$$

(b) Clearly there is only one way to match all the sandwiches correctly, so $\Pr(X = r \mid r = n) = 1/n!$. Also, note that it is impossible to match all but one sandwich, so $\Pr(X = r \mid r = n-1) = 0$. Only the cases for $r \leq n-2$ are nontrivial. Using a similar argument as part (a), we see that for any set of m sandwiches, the probability that at least one was correctly distributed is

$$\Pr\left(\bigcup_{k=1}^{m} A_k\right) = \sum_{k=1}^{m} (-1)^{k-1} \frac{(n-k)!}{n!}$$

and that the probability that *any* set of m sandwiches contained at least one correct match is

$$\sum_{k=1}^{m} (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{m} (-1)^{k-1} \frac{n!}{(n-k)!k!} \frac{(n-k)!}{n!}$$

$$= \sum_{k=1}^{m} \frac{(-1)^{k-1}}{k!} = 1 + \sum_{k=2}^{m} \frac{(-1)^{k-1}}{k!} = 1 - \sum_{k=2}^{m} \frac{(-1)^k}{k!}$$

So for $m \geq 2$, the probability of *no* correct matches is $\sum_{k=2}^{m} \frac{(-1)^k}{k!}$ if $n \geq 2$, and of course 0 if $n = 1$. Therefore the probability of r matches is the probability of any one set of r sandwiches all matching and none of the remaining $n-r$ sandwiches matching times the number of sets of r sandwiches; that is,

$$\Pr(X = r \mid r \leq n-2) = \binom{n}{r} \cdot \frac{(n-r)!}{n!} \cdot \left(\sum_{k=2}^{n-r} \frac{(-1)^k}{k!}\right) = \frac{r!}{(n-r)!r!} \cdot \frac{(n-r)!}{n!} \sum_{k=2}^{n-r} \frac{(-1)^k}{k!}$$

$$= \frac{1}{r!} \sum_{k=2}^{n-r} \frac{(-1)^k}{k!}$$

Therefore we have

$$\boxed{\Pr(X = r) = \begin{cases} \frac{1}{r!}\sum_{k=2}^{n-r}\frac{(-1)^k}{k!} & r \leq n-2 \\ 0 & r = n-1 \\ \frac{1}{r!} & r = n \end{cases}}$$

(c)

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{k=1}^{n} A_k\right) = \sum_{k=1}^{n}\mathbb{E}(A_k) = \sum_{k=1}^{n}\Pr(A_k = 1) = n \cdot \frac{1}{n} = \boxed{1}$$

(d)

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$\mathbb{E}(X^2) = \mathbb{E}\left(\sum_{k=1}^{n} A_k\right)^2 = \mathbb{E}\left(\sum_{k=1}^{n} A_k^2 + 2\sum_{1\leq i<j\leq n} A_i A_j\right) = \sum_{k=1}^{n}\mathbb{E}(A_k^2) + 2\sum_{1\leq i<j\leq n}\mathbb{E}(A_i A_j)$$

Because

$$\mathbb{E}(A_k^2) = 1^2 \cdot \Pr(A_k = 1) = \frac{1}{n}$$

$$\mathbb{E}(A_i A_j) = \Pr(A_j = 1 \cap A_j = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} = \frac{1}{2}\cdot\binom{n}{2}^{-1}$$

we have

$$\mathbb{E}(X^2) = \sum_{k=1}^{n}\frac{1}{n} + 2\sum_{1\leq i<j\leq n}\frac{1}{n(n-1)} = 1 + 2\cdot\binom{n}{2}\cdot\frac{1}{2}\cdot\binom{n}{2}^{-1} = 2$$

$$\implies \mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 2 - 1 = \boxed{1}$$

HW3 Problem 2(5). Verify: $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$ if X and Y are independent.

Solution. X and Y are independent if and only if

$$\Pr(X \cap Y) = \Pr(X)\cdot\Pr(Y) \iff \Pr(X = x \cap Y = y) = \Pr(X = x)\Pr(Y = y)$$

$$\iff \Pr(X = x \mid Y = y)\cdot\Pr(Y = y) = \Pr(X = x)\Pr(Y = y) \iff \Pr(X = x \mid Y = y) = \Pr(X = x)$$

$$\implies E(X \mid Y) = \sum_x x\cdot\Pr(X = x \mid Y = y) = \sum_x x\cdot\Pr(X = x) = \mathbb{E}(X)$$

HW3 Problem 2 (parts 1 - 4). Verify:

(1) $\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X)$

(2) $\mathbb{E}(g(Y)X \mid Y) = g(Y)\mathbb{E}(X \mid Y)$

(3) $\text{Cov}(\mathbb{E}(X \mid Y), Y) = \text{Cov}(X, Y)$

(4) Y and $X - \mathbb{E}(X \mid Y)$ are uncorrelated.

Solution.

(1)

$$\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) = \sum_y \mathbb{E}(X \mid Y)\Pr(Y = y) = \sum_y \left[\sum_x x \cdot \Pr(X = x \mid Y = y)\Pr(Y = y)\right]$$

$$= \sum_y \left[\sum_x x \cdot \Pr(X = x \cap Y = y)\right] = \sum_y \left[\sum_x x \cdot \Pr(Y = y \mid X = x) \cdot \Pr(X = x)\right]$$

$$= \sum_x \left[x \cdot \Pr(X = x) \cdot \sum_y \big(\Pr(Y = y \mid X = x)\big)\right] = \sum_x \left[x \cdot \Pr(X = x) \cdot 1\right]$$

$$= \mathbb{E}(X)$$

(2) 2

(3)

$$\text{Cov}(\mathbb{E}(X \mid Y), Y) = \mathbb{E}\Bigg(\Big[\mathbb{E}(X \mid Y) - \mathbb{E}\big(\mathbb{E}(X \mid Y)\big)\Big]\Big[Y - \mathbb{E}(Y)\Big]\Bigg)$$

$$= \mathbb{E}\Bigg(\Big[\mathbb{E}(X \mid Y) - \mathbb{E}(X)\Big]\Big[Y - \mathbb{E}(Y)\Big]\Bigg) = \mathbb{E}\Bigg(\mathbb{E}(X \mid Y)Y - \mathbb{E}(X)Y - \mathbb{E}(X \mid Y)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)\Bigg)$$

$$= \mathbb{E}\big(\mathbb{E}(X \mid Y)Y\big) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}\big(\mathbb{E}(X \mid Y)Y\big) - \mathbb{E}(Y)\mathbb{E}(X)$$

$$= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \text{Cov}(X, Y)$$

(4) Y and $X - \mathbb{E}(X \mid Y)$ are uncorrelated if and only if $\text{Cov}\big(Y, X - \mathbb{E}(X \mid Y)\big) = 0 \iff \mathbb{E}(Y \cdot \big[X - \mathbb{E}(X \mid Y)\big]) - \mathbb{E}(Y)\mathbb{E}(X - \mathbb{E}(X \mid Y)) = 0$.

$$\mathbb{E}(Y \cdot \big[X - \mathbb{E}(X \mid Y)\big]) - \mathbb{E}(Y)\mathbb{E}(X - \mathbb{E}(X \mid Y)) = \mathbb{E}(YX - Y\mathbb{E}(X \mid Y)) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(Y)\mathbb{E}\big(\mathbb{E}(X \mid Y)\big))$$

$$= \mathbb{E}(YX) - \mathbb{E}\big(Y\mathbb{E}(X \mid Y)\big) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(Y)\mathbb{E}(X) = \mathbb{E}(YX) - \mathbb{E}(YX) = 0$$

Spring 2018 Problem 2 (did not complete)

2. Consider positions 1 to n arranged in a circle, so that 2 comes after 1, 3 comes after 2, ..., n comes after $n-1$, and 1 comes after n. Similarly, take 1 to n as values, with cyclic order, and consider all $n!$ ways to assign values to positions, bijectively, with all $n!$ possibilities equally likely. For $i = 1$ to n, let X_i be the indicator that position i and the one following are filled in with two consecutive values in increasing order, and define

$$S_n = \sum_{i=1}^{n} X_i, \quad T_n = \sum_{i=1}^{n} iX_i$$

For example, with $n = 6$ and the the circular arrangement 314562, we get $X_3 = 1$ since 45 are consecutive in increasing order, and similarly $X_4 = X_6 = 1$, so that $S_6 = 3, T_6 = 13$.

a) Compute the mean and the variance of S_n.

b) Compute the mean and the variance of T_n.

Fall 2008 Problem 2 (HW1 Problem 10). Consider a lottery with n^2, tickets, of which only n tickets win prizes. Let p_n be the probability that, out of n randomly selected tickets, at least one wins a prize. Compute $\lim_{n\to\infty} p_n$.

Solution. There are $\binom{n^2}{n}$ possible sets of n tickets. The number of these sets that do not contain at least one winner (that is, they only contain members of the $n^2 - n$ losing tickets) is $\binom{n^2-n}{n}$. Therefore the probability of selecting a set of n tickets that contains at least one winner is

$$p_n = 1 - \binom{n^2-n}{n} \bigg/ \binom{n^2}{n} = 1 - \frac{(n^2-n)!}{n!(n^2-n-n)!} \bigg/ \frac{(n^2)!}{(n^2-n)!n!} = 1 - \frac{(n^2-n)!}{n!(n^2-2n)!} \cdot \frac{(n^2-n)!n!}{(n^2)!}$$

$$= 1 - \frac{(n^2-n)!}{(n^2-2n)!} \cdot \frac{(n^2-n)!}{(n^2)!} = 1 - \prod_{i=0}^{n-1} (n^2-n-i) \bigg/ \prod_{i=0}^{n-1} (n^2-i) = 1 - \prod_{i=0}^{n-1} \frac{n^2-n-i}{n^2-i}$$

$$= 1 - \prod_{i=0}^{n-1} \left(\frac{n^2-i}{n^2-i} - \frac{n}{n^2-i} \right) = 1 - \prod_{i=0}^{n-1} \left(1 - \frac{n}{n^2-i} \right)$$

Therefore

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} \left[1 - \prod_{i=0}^{n-1} \left(1 - \frac{n}{n^2-i} \right) \right] = 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{n}{n^2-i} \right) = 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{n \cdot \frac{1}{n}}{\frac{n^2}{n} - \frac{i}{n}} \right)$$

$$= 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{1}{n - \frac{i}{n}} \right) = 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{1}{n} \right) = 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n} \right)^n = \boxed{1 - \exp(-1)}$$

1.2.3 Problems we did on homework

Fall 2017 Problem 2 (Homework 3 Problem 6). An urn contains $2n$ balls, coming in pairs: two balls are labeled "1", two balls are labeled "2", ..., two balls are labeled "n". A sample of size n is taken without replacement. Denote by N the number of pairs in the sample. Compute the expected value and the variance of N. **You do not need to simplify the expression for the variance.**

Solution. Let X_k be an indicator variable for both balls labeled k being in the sample. Note that

$$\mathbb{E}(X_k) = \Pr(X_k = 1) = \frac{\binom{2n-2}{n-2}}{\binom{2n}{n}} = \frac{(2n-2)!}{(n-2)!n!} \Big/ \frac{(2n)!}{n!n!} = \frac{(2n-2)!n!}{(2n)!(n-2)!} = \frac{n(n-1)}{2n(2n-1)} = \frac{n-1}{2(2n-1)}$$

Now since $N = \sum_{k=1}^{n} X_k$, we have

$$\mathbb{E}(N) = \mathbb{E}\left(\sum_{k=1}^{n} X_k \right) = \sum_{k=1}^{n} \mathbb{E}(X_k) = \boxed{\frac{n(n-1)}{2(2n-1)}}$$

To obtain the variance, note that

$$\mathbb{E}(N^2) = \mathbb{E}\left(\sum_{k=1}^{n} X_k \right)^2 = \mathbb{E}\left(\sum_{k=1}^{n} X_k^2 + 2 \sum_{1 \le i < j \le n} X_i X_j \right) = \sum_{k=1}^{n} \mathbb{E}(X_k^2) + 2 \sum_{1 \le i < j \le n} \mathbb{E}(X_i X_j)$$

Because

$$\mathbb{E}(X_k^2) = 1^2 \cdot \Pr(X_k = 1) = \mathbb{E}(X_k) = \frac{n-1}{2(2n-1)}$$

$$\mathbb{E}(X_i X_j) = \Pr(X_i = 1 \cap X_j = 1) = \frac{\binom{2n-4}{n-4}}{\binom{2n}{n}} = \frac{(2n-4)!}{(n-4)!n!} \Big/ \frac{(2n)!}{n!n!} = \frac{(2n-4)!n!}{(2n)!(n-4)!}$$

$$= \frac{n(n-1)(n-2)(n-3)}{2n(2n-1)(2n-2)(2n-3)} = \frac{(n-1)(n-2)(n-3)}{2(2n-1)(2n-2)(2n-3)}$$

we have

$$\mathbb{E}(N^2) = \sum_{k=1}^{n} \frac{n-1}{2(2n-1)} + 2 \sum_{1 \le i < j \le n} \frac{(n-1)(n-2)(n-3)}{2(2n-1)(2n-2)(2n-3)} = \frac{n(n-1)}{2(2n-1)} + 2 \binom{n}{2} \frac{(2n-4)!n!}{(2n)!(n-4)!}$$

$$= \frac{n(n-1)}{2(2n-1)} + \frac{n!}{(n-2)!} \cdot \frac{(2n-4)!n!}{(2n)!(n-4)!} = \frac{n(n-1)}{2(2n-1)} + n(n-1) \cdot \frac{(n-1)(n-2)(n-3)}{2(2n-1)(2n-2)(2n-3)}$$

$$= \frac{n(n-1)}{2(2n-1)} + \frac{n(n-1)^2(n-2)(n-3)}{2(2n-1)(2n-2)(2n-3)}$$

$$\implies \operatorname{Var}(N) = \mathbb{E}(N^2) - \mathbb{E}(N)^2 = \boxed{\frac{n(n-1)}{2(2n-1)} + \frac{n(n-1)^2(n-2)(n-3)}{2(2n-1)(2n-2)(2n-3)} - \frac{n^2(n-1)^2}{4(2n-1)^2}}$$

Fall 2017 Problem 3 (HW3 Problem 8—almost full solution)

Let $U_1, U_2, \ldots$ be iid random variables, uniformly distributed on $[0, 1]$, and let N be a Poisson random variable with mean value equal to 1. Assume that N is independent of $U_1, U_2, \ldots$ and define

$$Y = \begin{cases} 0 & \text{if } N = 0 \\ \max\limits_{1 \le i \le N} U_i & \text{if } N > 0 \end{cases}$$

Compute the expected value of Y.

Solution. Since Y is a function of N, let $Y = y(N)$. By the Law of the Unconscious Statistician,

$$\mathbb{E}(Y) = \mathbb{E}\big(\mathbb{E}(Y \mid N)\big) = \mathbb{E}\big(\mathbb{E}(\max_{1 \le i \le N} U_i \mid N = n)\big)$$

Let $Z_n = \max_{1 \le i \le n} U_i$. The cdf of Z_n can be calculated as follows:

$$\Pr(Z_n \le x) = \Pr(\max_{1 \le i \le n} U_i \le x) = \Pr(U_1 \le x \cap U_2 \le x \cap \ldots \cap U_n \le x) = x^n$$

for $x \in [0, 1]$. Therefore the pdf of Z_n is its derivative, nx^{n-1}. So we have

$$\mathbb{E}(\max_{1 \le i \le N} U_i \mid N = n) = \mathbb{E}(Z_n) = \int_0^1 xnx^{n-1} dx = n \int_0^1 x^n dx = n \frac{x^{n+1}}{n+1} \bigg|_0^1 = \frac{n}{n+1}$$

Plugging this into the expression for $\mathbb{E}(Y)$ yields

$$\mathbb{E}(Y) = \mathbb{E}\big(\mathbb{E}(Y \mid N)\big) = \sum_{n=0}^{\infty} \frac{n}{n+1} \Pr(N = n) = \sum_{n=1}^{\infty} \frac{n}{n+1} \frac{\exp(-1)1^n}{n!}$$

$$= \frac{1}{e} \sum_{n=1}^{\infty} \frac{n+1-1}{(n+1)!} = \frac{1}{e}\left(\sum_{n=1}^{\infty} \frac{n+1}{(n+1)!} - \sum_{n=1}^{\infty} \frac{1}{(n+1)!} \right) = \frac{1}{e}\left(\sum_{n=1}^{\infty} \frac{1}{n!} - \sum_{m=2}^{\infty} \frac{1}{m!} \right)$$

$$= \frac{1}{e}\big[e - 1 - (e - 1 - 1)\big] = \boxed{\frac{1}{e}}$$

Fall 2013 Problem 3/Spring 2011 Problem 2 (HW3 Problem 9; coupon collector problem) Only parts I didn't do: 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 $\Pr(D) \to a$?

HW3 Problem 9. Consider n (different) balls placed at random in m boxes so that each of m^n configurations is equally likely.

(a) Compute the expected value and the variance of the number of empty boxes.

(b) Show that if $\lim_{m,n\to\infty} m\exp(-n/m) = \lambda \in (0,\infty)$, then, in the same limit, the number of empty boxes has Poisson distribution with parameter λ.

(c) For $k \geq 1$ such that $k+3 \leq m$, define the event A_k that the boxes $k, k+1, k+2, k+3$ are empty. Assuming that $m > 8$, compute $\Pr(A_1 \cup A_3 \cup A_5)$. How will the answer change if $m = 8$?

(d) Now imagine that the balls are dropped one-by-one (with each ball equally likely to go into any of the m boxes, independent of all other balls), and denote by N_m the minimal number of balls required to fill all the boxes. Compute $\mathbb{E}(N_m), \mathrm{Var}(N_m)$ and

$$\lim_{m\to\infty} \Pr\left(\frac{N_m - m\log m}{m} \leq x\right)$$

(e) Suppose we instead place an unlimited number of balls into the m boxes until we have k consecutive balls land in the same box (it doesn't matter which box). What is the expected number of balls we will drop until this happens?

Solution.

(a) Let A_i be the event that the ith box is empty. Let $\mathbf{1}_{A_i}$ be the indicator for A_i. Then $X = \sum_{i=1}^m \mathbf{1}_{A_i}$.

$$\mathbb{E}(X) = \mathbb{E}\bigg(\sum_{i=1}^m \mathbf{1}_{A_i}\bigg) = \sum_{i=1}^m \left(\mathbb{E}\mathbf{1}_{A_i}\right) = \sum_{i=1}^m \Pr(A_i) = \sum_{i=1}^m \left(\frac{m-1}{m}\right)^n = \boxed{\frac{(m-1)^n}{m^{n-1}}}$$

$$\mathrm{Var}(X) = \mathrm{Var}\bigg(\sum_{i=1}^m \mathbf{1}_{A_i}\bigg) = \sum_{i=1}^m \mathrm{Var}(\mathbf{1}_{A_i}) + 2\sum_{1\leq i<j\leq m} \mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j})$$

$$\mathrm{Var}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) = \mathbb{E}(\mathbf{1}_{A_i}\mathbf{1}_{A_i}) - \mathbb{E}(\mathbf{1}_{A_i})^2 = \Pr(A_i \cap A_i) - \Pr(A_i)^2 = \left(\frac{m-1}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}$$

$$\mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) = \mathbb{E}(\mathbf{1}_{A_i}\mathbf{1}_{A_j}) - \mathbb{E}(\mathbf{1}_{A_i})\mathbb{E}(\mathbf{1}_{A_j}) = \Pr(A_i \cap A_j) - \Pr(A_i)\Pr(A_j) = \left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}$$

$$\implies \mathrm{Var}(X) = m \cdot \left[\left(\frac{m-1}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right] + \frac{m!}{(m-2)!}\left[\left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right]$$

$$= \frac{(m-1)^n}{m^{n-1}} - \frac{(m-1)^{2n}}{m^{2n-1}} + (m^2 - m)\left[\left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right]$$

$$\boxed{\mathrm{Var}(X) = \frac{(m-1)^n}{m^{n-1}} - \frac{(m-1)^{2n}}{m^{2n-1}} + (m-1)\left[\frac{(m-2)^n}{m^{n-1}} - \frac{(m-1)^{2n}}{m^{2n-1}}\right]}$$

(b) Note that

$$X = \sum_{i=1}^m \mathbf{1}_{A_i}$$

and that the A_i are only weakly dependent on each other, especially as m and n increase. Therefore as $m, n \to \infty$, the Poisson paradigm (see Section 1.1.8) suggests $X \sim \mathrm{Poisson}(\mathbb{E}(X))$. We have

$$\mathbb{E}(X) = \frac{(m-1)^n}{m^{n-1}}$$

so

$$\lim_{n,m\to\infty} \mathbb{E}(X) = \lim_{n,m\to\infty} m \cdot \left(\frac{m-1}{m}\right)^n = \lim_{n,m\to\infty} m \cdot \left(1 - \frac{1}{m}\right)^n = \lim_{n,m\to\infty} m \cdot \left[\left(1-\frac{1}{m}\right)^m\right]^{n/m}$$

$$\approx \lim_{n,m\to\infty} m \cdot \left[e^{-1}\right]^{n/m} = \lim_{n,m\to\infty} me^{-n/m}$$

Using

$$\lim_{m,n\to\infty} m \exp(-n/m) = \lambda \in (0,\infty)$$

we have $\boxed{X \sim \text{Poisson}(\lambda) \text{ as } m,n \to \infty}$.

(c)

$$\Pr(A_1 \cup A_3 \cup A_5) = \Pr(A_1) + \Pr(A_3) + \Pr(A_5) - \Pr(A_1 \cap A_3) - \Pr(A_1 \cap A_5) - \Pr(A_3 \cap A_5) + \Pr(A_1 \cap A_3 \cap A_5)$$

We have

$$\Pr(A_1) = \Pr(A_3) = \Pr(A_5) = \left(\frac{m-4}{m}\right)^n$$

$$\Pr(A_1 \cap A_3) = \Pr(A_3 \cap A_5) = \left(\frac{m-6}{m}\right)^n$$

$$\Pr(A_1 \cap A_5) = \Pr(A_1 \cap A_3 \cap A_5) = \left(\frac{m-8}{m}\right)^n$$

Therefore

$$\Pr(A_1 \cup A_3 \cup A_5) = 3\left(\frac{m-4}{m}\right)^n - 2\left(\frac{m-6}{m}\right)^n = \boxed{\frac{3(m-4)^n - 2(m-6)^n}{m^n}}$$

(d) N_m is the minimal number of balls required to fill all the boxes. Let T_i be the number of balls that have to be dropped to fill the ith box after $i-1$ boxes have been filled. The probability of filling a new box after $i-1$ boxes have been filled is $\frac{m-(i-1)}{m}$. Therefore T_i has a geometric distribution with $E(T_i) = \frac{m}{m-(i-1)}$. Since $N_m = \sum_{i=1}^m T_i$, we have

$$\mathbb{E}(N_m) = \mathbb{E}\left(\sum_{i=1}^m T_i\right) = \sum_{i=1}^m \mathbb{E}(T_i) = \sum_{i=1}^m \frac{m}{m-(i-1)} = \boxed{m \sum_{i=1}^m \frac{1}{i}}$$

Because the T_i are independent, we have

$$\text{Var}(N_m) = \text{Var}\left(\sum_{i=1}^m T_i\right) = \sum_{i=1}^m \text{Var}(T_i) = \sum_{i=1}^m \left(1 - \frac{m-(i-1)}{m}\right) \Big/ \left(\frac{m-(i-1)}{m}\right)^2$$

$$= \sum_{i=1}^{m} \frac{i-1}{m} \cdot \left(\frac{m}{m-(i-1)} \right)^2 = \boxed{m \sum_{i=1}^{m} \frac{i-1}{[m-(i-1)]^2}}$$

Finally, to find

$$\lim_{m \to \infty} \Pr \left(\frac{N_m - m \log m}{m} \leq x \right)$$

begin by noting that we can also express N_m as

$$\Pr(N_m \leq k) = \Pr\left(X_{m,k} = 0\right)$$

where $X_{m,k}$ is defined as X is in part (b) with k being the number of balls that have been dropped so far, $k \in \mathbb{N} \geq m$. (For $k < m$, $\Pr(N_m \leq k) = 0$.)

Again, let $A_{i,k}$ be the event that the ith box is empty after dropping k balls. Then because $X_{m,k} = \sum_{i=1}^{m} \mathbf{1}_{A_{i,k}}$ and the $A_{i,k}$ are only weakly dependent on each other (especially as m becomes large), the Poisson paradigm (see Section 1.1.8) again suggests that as $m \to \infty$, $X_{m,k} \sim \text{Poisson}(\lambda_k)$ where $\lambda_k = \mathbb{E}(X_{m,k})$ is defined as above. Therefore we have

$$\lim_{m \to \infty} \Pr \left(\frac{N_m - m \log m}{m} \leq x \right) = \lim_{m \to \infty} \Pr\left(N_m \leq xm + m \log m\right) = \lim_{m \to \infty} \Pr\left(X_{m, xm + m \log m}\right.$$

$$\left. = 0\right) \approx \frac{\exp(-\lambda_{xm+m\log m}) \cdot \lambda_{xm+m\log m}^0}{0!} = \exp\left(- \lambda_{xm+m\log m}\right)$$

And we have

$$\lambda_{xm+m\log m} = \lim_{m \to \infty} m \exp\left(- \frac{xm + m \log m}{m} \right) = \lim_{m \to \infty} m \exp\left(- x - \log m\right) = \lim_{m \to \infty} m/m \exp(-x)$$

$$= \exp(-x)$$

which yields

$$\boxed{\lim_{m \to \infty} \Pr \left(\frac{N_m - m \log m}{m} \leq x \right) = \exp(\exp(-x))}$$

(e) Let $N = N_k$ be the number of balls that are dropped until k consecutive balls land in the same box, and likewise for N_{k-1}. Suppose we have already observed $k-1$ consecutive outcomes (of any kind) in N_{k-1} trials. Then we finish on the next term (by having another consecutive outcome) with probability $1/m$. Otherwise we have a different outcome and then repeat the same process again. So we have

$$\mathbb{E}(N_k \mid N_{k-1}) = N_{k-1} + 1 \cdot \frac{1}{m} + \mathbb{E}(N_k) \cdot \left(1 - \frac{1}{m}\right)$$

Therefore

$$\mathbb{E}(N) = \mathbb{E}(N_k) = \mathbb{E}\big[\mathbb{E}(N_k \mid N_{k-1})\big] = \mathbb{E}(N_{k-1}) + \frac{1}{m} + \left(1 - \frac{1}{m}\right)\mathbb{E}(N_k)$$

$$\iff \frac{1}{m}\mathbb{E}(N_k) = \mathbb{E}(N_{k-1}) + \frac{1}{m} \iff \mathbb{E}(N_k) = m\mathbb{E}(N_{k-1}) + 1$$

We have a recursive formula. Note that $\mathbb{E}(N_1) = 1$ because the number of trials until there is 1 consecutive outcome of any kind is simply 1. We can then calculate as follows:

$$\mathbb{E}(N_2) = m\mathbb{E}(N_{2-1}) + 1 = m + 1$$

$$\mathbb{E}(N_3) = m\mathbb{E}(N_{3-1}) + 1 = m(m+1) + 1 = 1 + m + m^2$$

$$\mathbb{E}(N_4) = m\mathbb{E}(N_{4-1}) + 1 = m(1 + m + m^2) + 1 = 1 + m + m^2 + m^3$$

$$\vdots$$

$$\mathbb{E}(N_k) = \sum_{i=0}^{k-1} m^i = \frac{1 \cdot (1 - m^k)}{1 - m} = \boxed{\frac{m^k - 1}{m - 1}}$$

Fall 2012 Problem 1 (HW2 Problem 10/HW 1 Problem 9) Only part I didn't do: Find the mean and variance of $S_n = X_1 + \ldots + X_n$, the total number of white balls added to the urn up to time n.

HW1 Problem 9. An urn contains b black and w white balls. At each step, a ball is removed from the urn at random and then put back together with one more ball of the same color. Compute the probability p_n to get a black ball on step $n, n \geq 1$.

Solution. Step 1:

$$p_1 = \frac{b}{b+w}$$

Step 2: We need to separately consider the cases where a black ball was selected on step 1 (with probability p_1) or a white ball (with probability $1 - p_1$).

$$p_2 = p_1 \cdot \frac{b+1}{b+w+1} + (1 - p_1) \cdot \frac{b}{b+w+1} = p_1\left(\frac{b+1}{b+w+1} - \frac{b}{b+w+1}\right) + \frac{b}{b+w+1}$$

$$= p_1\left(\frac{1}{b+w+1} + \frac{1}{p_1}\frac{b}{b+w+1}\right) = p_1\left(\frac{1}{b+w+1} + \frac{b+w}{b}\frac{b}{b+w+1}\right)$$

$$= p_1\left(\frac{b+w+1}{b+w+1}\right) = p_1$$

$$\implies p_2 = p_1 = \frac{b}{b+w}$$

Step 3: Regardless of the previous steps, there are now $b + w + 2$ balls in the urn. Since we know that $p_1 = p_2$, the probability that we have selected k black balls so far (and thus, the probability that there are currently $b + k$ black balls in the urn) is given by

$$\Pr(k \text{ balls chosen in first 2 rounds}) = \binom{2}{k} p_1^k (1 - p_1)^{2-k} = \binom{2}{k}\left(\frac{b}{b+w}\right)^k \left(\frac{w}{b+w}\right)^{2-k}$$

$$= \binom{2}{k}\frac{b^k w^{2-k}}{(b+w)^2}$$

for $k \in \{0, 1, 2\}$. Given that we have selected k black balls so far, the probability of selecting a black ball this time is $\frac{b+k}{b+w+2}$. Therefore the probability of selecting a black ball this round is

$$p_3 = \sum_{k=0}^{2}\binom{2}{k}\frac{b^k w^{2-k}}{(b+w)^2}\frac{b+k}{b+w+2} = \frac{1}{(b+w+2)(b+w)^2}\sum_{k=0}^{2}\binom{2}{k}(b+k)b^k w^{2-k}$$

$$= \frac{1}{(b+w+2)(b+w)^2}\left(\binom{2}{0}bw^2 + \binom{2}{1}(b+1)bw + \binom{2}{2}(b+2)b^2\right)$$

$$= \frac{bw^2 + 2(b+1)bw + (b+2)b^2}{(b+w+2)(b+w)^2} = \frac{b}{b+w}\left(\frac{w^2 + 2bw + 2w + b^2 + 2b}{b^2 + bw + 2b + wb + w^2 + 2w}\right)$$

$$= \frac{b}{b+w}\left(\frac{w^2 + 2bw + 2w + b^2 + 2b}{b^2 + 2bw + 2b + w^2 + 2w}\right) = \frac{b}{b+w} = p_1$$

There seems to be a clear pattern here. Let's find the general formula by induction.

Step $n + 1$: Assume that the probability of choosing a black ball on steps $1, 2, \ldots, n$ was $\frac{b}{b+w}$ each time.

(a bunch of boring stuff, then it worked.)

HW2 Problem 10. Random variables $(X_1, \ldots, X_n)$ are called *exchangeable* if $\Pr(X_1 = x_1, \ldots, X_n = x_n) = \Pr(X_{\tau(1)} = x_1, \ldots, X_{\tau(n)} = x_n)$ for all real numbers $x_1, \ldots, x_n$ and every permutation τ of the set $\{1, \ldots, n\}$. In the setting of Problem 9 from Homework 1, let $X_k = 1$ if a white ball is drawn on step k, and $X_k = 0$ otherwise. Show that the random variables $X_1, \ldots, X_n$ are exchangeable for every $n \geq 2$.

Solution. For $n = 2$: There are two cases which we must show are equal to show exchangeability:

$$\Pr(X_1 = 0, X_2 = 1) = \Pr(X_1 = 1, X_2 = 0)$$

First,

$$\Pr(X_1 = 0, X_2 = 1) = \Pr(\text{black first}) \Pr(\text{white second} \mid \text{black first}) = \left(\frac{b}{b+w} \right) \left(\frac{w}{b+w+1} \right)$$

$$\left(\frac{w}{b+w} \right) \left(\frac{b}{b+w+1} \right) = \Pr(X_1 = 1, X_2 = 0)$$

which proves exchangeability for $n = 2$. In the general case, we seek to show that $X_1, \ldots, X_n$ are exchangeable. That is, in all $n+1$ unordered sets $\mathbb{X}_k = \{x_{1k}, x_{2k}, \ldots, x_{nk} \mid x_{ik} \in \{0,1\}, \sum_i x_{ik} = k\}$, in all $\binom{n}{k}$ permutations of $\mathbb{X}_k$,

$$\Pr(\mathbb{X}_{kj} = \Pr(\mathbb{X}_{kj'}$$

where j and j' denote different permutations of $\mathbb{X}_k$. That is,

$$\Pr(X_1 = x_{1k}, X_2 = x_{2k}, \ldots, X_n = x_{nk}) = \Pr(X_{j_1} = x_{1k}, X_{j_2} = x_{2k}, \ldots, X_{j_n} = x_{nk})$$

where $j_1, j_2, \ldots, j_n$ index the permuted variables. Consider $\mathbb{X}_{kj^*}$ where all k white balls are chosen first and all $n-k$ black balls are chosen last. We have

$$\Pr(\mathbb{X}_{kj^*}) = \prod_{i=1}^{k} \left(\frac{w+i-1}{b+w+i-1} \right) \cdot \prod_{i=k+1}^{n} \left(\frac{b+i-k-1}{b+w+i-1} \right)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{b+w+i-1} \right) \cdot \left[\prod_{i=1}^{k} (w+i-1) \prod_{i=k+1}^{n} (b+i-k-1) \right] = \prod_{i=1}^{n} \left(\frac{1}{b+w+i-1} \right) \cdot \left[\prod_{i=1}^{k} (w+i-1) \prod_{i'=1}^{n-k} (b+i'-1) \right]$$

It is easy to see that the leftmost product will always equal the product of the denominators, regardless of the permutation, since one ball is added to the urn after every draw. Similarly, regardless of permutation, the numerator of the probability of drawing the ith white ball will always equal $w + i - 1$, the number of white balls already in the urn. Likewise, the numerator of the probability of drawing the i'th black ball is always $b + i' - 1$. Because multiplication is commutative, all permutations of these numbers will have equal products. Therefore $\Pr(\mathbb{X}_{kj^*}) = \Pr(\mathbb{X}_{kj}$ for all k. That is,

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \Pr(X_{\tau(1)} = x_1, \ldots, X_{\tau(n)} = x_n)$$

for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$, all $n \in \mathbb{Z}$ such that $n \geq 2$, all permutations τ.

Homework 2 Problem 2. Consider the function

$$f(x) = \begin{cases} C(2x - x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Could f be a distribution function? If so, determine C.

(b) Could f be a probability density function? If so, determine C.

Solution.

(a) If f is a distribution function, $\lim_{x\to-\infty} f(x) = 0$, $\lim_{x\to\infty} f(x) = 1$, and $f'(x) \geq 0 \ \forall \ x \in \mathbb{R}$. f clearly does not meet the second or third conditions and is therefore $\boxed{\text{not a distribution function.}}$

(b) If f is a density function then $\int_{-\infty}^{\infty} f(x)dx = 1$ and $f(x) \geq 0 \ \forall x \in \mathbb{R}$.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{2} C(2x - x^2)dx = C\left[x^2 - \frac{x^3}{3}\right]_0^2 = C\left(4 - \frac{8}{3} - 0\right) = C \cdot \frac{4}{3}$$

$$= 1 \iff C = \frac{3}{4}$$

Next we check that f is always nonnegative. It equals zero except on $(0, 2)$.

$$\frac{3}{4}(2x - x^2) \geq 0 \iff x(2 - x) \geq 0 \iff x \in (0, 2)$$

Therefore f is nonnegative $\forall x \in \mathbb{R}$, so $\boxed{f \text{ is a probability density function if } C = \frac{3}{4}.}$

HW1 Problem 8. Two people, A and B, are involved in a duel. The rules are simple: shoot at each other once; if at least one is hit, the duel is over, if both miss, repeat (go to the next round), and so on. Denote by p_A and p_B the probabilities that A hits B and B hits A with one shot, and assume that that hitting/missing is independent from round to round. Compute the probabilities of the following events: (a) the duel ends and A is not hit; (b) the duel ends and both are hit; (c) the duel ends after round number n; (d) the duel ends after round number n GIVEN that A is not hit; (e) the duel ends after n rounds GIVEN that both are hit; (f) the duel goes on forever.

Solution.

(a) Let A_k denote the event that the duel is ended by A shooting B in the kth round (with neither person being shot in the first $k - 1$ rounds). Note that $\{A_k | k = 1, 2, \ldots\}$ are all mutually exclusive. Therefore the probability of the duel ending without A being hit is $\sum_{k=1}^{\infty} A_k$. Because the probabilities in each round are constant and independent,

$$A_k = (1 - p_A)^{k-1} p_A (1 - p_B)^k$$

So the probability that the duel ends and A is not hit is

$$\sum_{k=1}^{\infty} A_k = \sum_{k=1}^{\infty} (1 - p_A)^{k-1} p_A (1 - p_B)^k = p_A(1 - p_B) \sum_{k=1}^{\infty} (1 - p_A)^{k-1}(1 - p_B)^{k-1}$$

This is an infinite geometric series. Since the ratio $(1 - p_A)(1 - p_B)$ has absolute value less than 1, the sum can be calculated.

$$\sum_{k=1}^{\infty} A_k = p_A(1 - p_B) \cdot \frac{1}{1 - (1 - p_A)(1 - p_B)} = \frac{p_A(1 - p_B)}{p_A + p_B - p_A p_B} = \boxed{\frac{p_A(1 - p_B)}{p_A(1 - p_B) + p_B}}$$

(b) Similar to part (a). Let C_k denote the event that the duel is ended with both players being shot in the kth round (with neither person being shot in the first $k-1$ rounds). Again, $\{C_k|k=1,2,\ldots\}$ are all mutually exclusive, so the probability of the duel ending in these circumstances is $\sum_{k=1}^{\infty} C_k$. We have

$$C_k = (1-p_A)^{k-1}p_A(1-p_B)^{k-1}p_B$$

$$\sum_{k=1}^{\infty} C_k = \sum_{k=1}^{\infty}(1-p_A)^{k-1}p_A(1-p_B)^{k-1}p_B = p_A p_B \sum_{k=1}^{\infty}(1-p_A)^{k-1}(1-p_B)^{k-1}$$

$$= p_A p_B \cdot \frac{1}{1-(1-p_A)(1-p_B)} = \boxed{\frac{p_A p_B}{p_A + p_B - p_A p_B}}$$

Note that this value is less than the answer from part (a) if $p_B < \frac{1}{2}$ and greater if $p_B > \frac{1}{2}$

(c) Let B_k denote the event that the duel is ended by B shooting A in the kth round (with neither person being shot in the first $k-1$ rounds), with

$$B_k = (1-p_A)^k p_B(1-p_B)^{k-1}$$

Let A_k and C_k be defined as above. Note that $\{A_k|k=1,2,\ldots\}, \{B_k|k=1,2,\ldots\}, \{C_k|k=1,2,\ldots\}$ are all mutually exclusive, and that the event that the duel ends in round n is $\{A_n \cup B_n \cup C_n\}$. So the probability of the duel ending in round n is

$$\Pr(A_n \cup B_n \cup C_n) = \Pr(A_n) + \Pr(B_n) + \Pr(C_n)$$

$$= (1-p_A)^{n-1}p_A(1-p_B)^n + (1-p_A)^n p_B(1-p_B)^{n-1} + (1-p_A)^{n-1}p_A(1-p_B)^{n-1}p_B$$

$$= (1-p_A)^{n-1}(1-p_B)^{n-1}\big[p_A(1-p_B) + (1-p_A)p_B + p_A p_B\big]$$

$$= \boxed{(1-p_A)^{n-1}(1-p_B)^{n-1}(p_A + p_B - p_A p_B)}$$

(d) Let A_k, B_k, C_k be defined as above. The event that the duel ends at round n without A being hit is given by $\{A_n\}$.

$$\Pr(A_n) = \boxed{(1-p_A)^{n-1}p_A(1-p_B)^n}$$

(e) Let A_k, B_k, C_k be defined as above. The event that the duel ends at round n with both players being hit is given by $\{C_n\}$.

$$\Pr(C_n) = \boxed{(1-p_A)^{n-1}p_A(1-p_B)^{n-1}p_B}$$

(f) Let A_k, B_k, C_k be defined as above. The probability that the duel never ends is equal to 1 - the probability that the duel ends at some point, which is $\{A_k|k=1,2,\ldots\} \cup \{B_k|k=1,2,\ldots\} \cup \{C_k|k=1,2,\ldots\}$. Since all of these events are mutually exclusive, we have

$$1 - \Pr(\{A_k|k=1,2,\ldots\} \cup \{B_k|k=1,2,\ldots\} \cup \{C_k|k=1,2,\ldots\}) = 1 - \sum_{k=1}^{\infty}\big(A_k + B_k + C_k\big)$$

$$= 1 - \sum_{k=1}^{\infty} \big((1-p_A)^{k-1} p_A (1-p_B)^k + (1-p_A)^k p_B (1-p_B)^{k-1} + (1-p_A)^{k-1} p_A (1-p_B)^{k-1} p_B\big)$$

$$= 1 - \big[p_A(1-p_B) + (1-p_A)p_B + p_A p_B\big] \sum_{k=1}^{\infty} (1-p_A)^{k-1}(1-p_B)^{k-1}$$

$$= 1 - \big[p_A 1 - p_A p_B) + p_B - p_A)p_B + p_A p_B\big] \cdot \frac{1}{1-(1-p_A)(1-p_B)}$$

$$= 1 - \frac{p_A - p_A p_B + p_B - p_A p_B + p_A p_B}{p_A + p_B - p_A p_B} = 1 - \frac{p_A + p_B - p_A p_B}{p_A + p_B - p_A p_B} = \boxed{0}$$

Homework 1 Problem 1.

(I) Seven different gifts are distributed among 10 children. How many different outcomes are possible if every child can receive (a) at most one gift, (b) at most two gifts, (c) any number of gifts?

(II) Answer the same questions if the gifts are identical (but the children are still different).

Solution.

(I) (a) $\binom{10}{7} 7! = \boxed{604,800}$

(b) Clearly all outcomes that satisfy part (I)(a) also satisfy these conditions, so we start with $\binom{10}{7} 7! = 604,800$ possible outcomes. In addition, the following outcomes are possible:

(i) **A set of 6 children receive gifts; one child receives two gifts.** There are $\binom{10}{6}$ ways to pick a group of 6 children to receive the gifts. Next, there are $\binom{6}{1} = 6$ ways to choose which child receives two gifts. Finally, there are $7!/2!$ unique ways to distribute the gifts among the children once a particular partition is chosen (since order matters for all of the gifts except for the two that are received by the same child).

(ii) **A set of 5 children receive gifts; two children receive two gifts.** There are $\binom{10}{5}$ ways to pick a group of 5 children to receive the gifts. Next, there are $\binom{5}{2}$ ways to choose which of these children receive one gift and which receive two. Finally, there are $7!/(2!2!)$ unique ways to distribute the gifts among the children once a particular partition is chosen (since order matters for all of the gifts except for the two batches of two gifts that are received by the same child).
(Note that without the restriction that a child can receive at most two gifts, another possibility is that 1 child could receive 3 gifts, but that wouldn't work in this case.)

(iii) **A set of 4 children receive gifts; three children each receive two gifts.** There are $\binom{10}{4}$ ways to pick a group of 4 children to receive the gifts. Next, there are $\binom{4}{3} = 4$ ways to choose which of these children receive one gift and which receive two. Finally, there are $7!/(2!2!2!)$ unique ways to distribute the gifts among the children once a particular partition is chosen (since order matters for all of the gifts except for the three batches of two gifts that are received by the same child).
(Again, there are other possibilities for 4 children to receive 7 gifts, but none that satisfy the condition that no child receives more than 2 gifts.)

Clearly each of these outcomes are mutually exclusive. Therefore the answer is

$$\binom{10}{7}7! + \binom{10}{6} \cdot \binom{6}{1} \cdot \frac{7!}{2!} + \binom{10}{5} \cdot \binom{5}{2} \cdot \frac{7!}{2!2!} + \binom{10}{4} \cdot \binom{4}{3} \cdot \frac{7!}{2!2!2!}$$

$$= 7! \cdot \left(\frac{10!}{3!} + \frac{10!}{6!4!} \cdot 6 \cdot \frac{1}{2} + \frac{10!}{5!5!} \cdot \frac{5!}{3!2!} \cdot \frac{1}{4} + \frac{10!}{4!6!} \cdot \frac{4!}{3!} \cdot \frac{1}{8} \right)$$

$$= 7!10! \cdot \left(\frac{1}{3!} + \frac{1}{6!4!} \cdot \frac{6}{2} + \frac{1}{5!} \cdot \frac{1}{3!2!} \cdot \frac{1}{4} + \frac{1}{6!3! \cdot 8} \right)$$

$$= \boxed{7,484,400}$$

(c) $10^7 = \boxed{10,000,000}$

(II) (a) $\binom{10}{7} = \boxed{120}$

(b) Clearly all outcomes that satisfy part (I)(a) also satisfy these conditions, so we start with $\binom{10}{7} =$ 120 possible outcomes. In addition, the following outcomes are possible:

(i) A set of 6 children receive gifts; one child receives two gifts (6 distinct ways this could happen for each set of 6 children).

(ii) A set of 5 children receive gifts; two children receive two gifts ($\binom{5}{2}$ distinct ways this could happen for each set of 5 children).

(iii) A set of 4 children receive gifts; three children each receive two gifts (4 distinct ways this could happen for each set of 4 children).

Clearly each of these outcomes are mutually exclusive. Therefore the answer is

$$\binom{10}{7} + \binom{10}{6} \cdot \binom{6}{7-6} + \binom{10}{5} \cdot \binom{5}{7-5} + \binom{10}{4} \cdot \binom{4}{7-4} = \boxed{4,740}$$

(c) By Proposition 11, the number of nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1 + x_2 + \ldots + x_r = n$$

is equal to $\binom{n+r-1}{r-1} = \binom{7+10-1}{10-1} = \boxed{11,440}$.

Homework 1 Problem 2.

(I) 20 different gifts are distributed among seven children. How many different outcomes are possible if every child can receive (a) at least one gift, (b) at least two gifts, (c) any number of gifts?

(II) Answer the same questions if the gifts are identical (but the children are still different).

(III) Now try to generalize problems (1) and (2).

Solution.

(I) (a) There are 7^{20} possible allocations of gifts if we have no restrictions. If one child doesn't get a gift, there are $\binom{7}{1}$ ways to choose which child that is and 6^{20} subsequent allocations of gifts. Likewise, there are $\binom{7}{2} \cdot (7-2)^{20}$ ways to allocate the gifts if two children don't receive gifts, $\binom{7}{3} \cdot (7-3)^{20}$ ways if three children don't receive gifts, $\binom{7}{4} \cdot (7-4)^{20}$ ways if four children don't receive gifts, $\binom{7}{5} \cdot (7-5)^{20}$ ways if five children don't receive gifts, and $\binom{7}{6} \cdot (7-6)^{20}$ ways if six children don't receive gifts.

Let A_i denote the number of allocations in which i children do not receive gifts. In order to make the calculation, we must use the Inclusion-Exclusion principle (Proposition 5) (because, for example, some of the allocations in which three children don't receive gifts include allocations where four or more children don't receive gifts, and we don't want to double-count). Therefore the number of ways that at least one child can not receive a gift (i.e. the complement of every child receiving at least one gift) is

$$\left|\bigcup_{i=1}^{6} A_i\right| = \sum_{i=1}^{6} |A_i| - \sum_{1 \le i < j \le 6} |A_i \cap A_j| + \sum_{1 \le i < j < k \le 6} |A_i \cap A_j \cap A_k| - \dots$$

$$+(-1)^{6-1} |A_1 \cap A_2 \cap A_3 \cap \dots \cap A_6|$$

Fortunately, these allocations are nested in the sense that all the allocations where e.g. 5 children do not receive gifts are a subset of all the allocations where 4 children do not receive gifts; that is

$$A_6 \subset A_5 \subset A_4 \subset A_3 \subset A_2 \subset A_1$$

which implies e.g.

$$A_1 \cap A_2 \cap A_3 \cap \dots \cap A_6 = A_6,$$

$$\sum_{1 \le i < j \le 6} |A_i \cap A_j| = 5|A_6| + 4|A_5| + 3|A_4| + 2|A_3| + |A_2|$$

So we have

$$\left|\bigcup_{i=1}^{6} A_i\right| = |A_6| + |A_5| + |A_4| + |A_3| + |A_2| + |A_1| - (5|A_6| + 4|A_5| + 3|A_4| + 2|A_3| + |A_2|)$$

$$+(4|A_6| + 3|A_5| + 2|A_4| + |A_3|) - (3|A_6| + 2|A_5| + |A_4|) + \dots - |A_6|$$

$$= |A_1| - |A_2| + |A_3| - |A_4| + |A_5| - |A_6|$$

$$= \binom{7}{1} \cdot 6^{20} - \binom{7}{2} \cdot (7-2)^{20} + \binom{7}{3} \cdot (7-3)^{20} - \binom{7}{4} \cdot (7-4)^{20}$$

$$+\binom{7}{5} \cdot (7-5)^{20} - \binom{7}{6} \cdot (7-6)^{20}$$

The final answer is

$$7^{20} - \left|\bigcup_{i=1}^{6} A_i\right| = 7^{20} - \binom{7}{1}\cdot 6^{20} + \binom{7}{2}\cdot(7-2)^{20} - \binom{7}{3}\cdot(7-3)^{20} + \binom{7}{4}\cdot(7-4)^{20}$$

$$-\binom{7}{5}\cdot(7-5)^{20} + \binom{7}{6}\cdot(7-6)^{20} \approx \boxed{5.616\cdot 10^{16}}$$

(b) Similar to above, but more complicated. The complement of every child receiving at least two gifts is that at least one child doesn't receive a gift (same as above) or at least one child only receives one gift. So we start from the baseline answer above, and subtract out all the possible allocations in which at least one child receives one gift.

If one child only receives one gift (and the rest receive more than one), there are $\binom{7}{1}$ ways to choose which child that is, $\binom{20}{1}$ ways to choose which gift that child receives, and 6^{20-1} allocations of the remaining gifts. If two children receive only one gift, there are $\binom{7}{2}$ ways to choose which children those are, $\binom{20}{2}\cdot 2!$ ways to choose which gifts those children get and distribute them among those children, and $(7-2)^{20-2}$ ways to allocate the remaining gifts. Likewise, if three children receive only one gift there are $\binom{7}{3}\binom{20}{3}\cdot 3!\cdot(7-3)^{20-3}$ ways to allocate the gifts, $\binom{7}{4}\binom{20}{4}\cdot 4!\cdot(7-4)^{20-4}$ ways if four children receive only one gift, $\binom{7}{5}\binom{20}{5}\cdot 5!\cdot(7-5)^{20-5}$ ways if five children receive only one gift, and $\binom{7}{6}\binom{20}{6}\cdot 6!(7-6)^{20-6}$ ways if six children don't receive gifts.

Let B_j be the event that j children receive only one gift. Note that $B_1 \cap A_i$ is nonempty $\forall i < 7-1$, $B_2 \cap A_i$ is nonempty $\forall i < 7-2$, and in general, $B_j \cap A_i$ is nonempty $\forall i < 7-j, j \in \{1,2,\ldots,6\}$. Applying the Inclusion-Exclusion Principle (Proposition 5) in a similar way as in part (I)(a), the answer is

$$\boxed{7^{20} - \left|\bigcup_{i=1}^{6} A_i\right| - \left|\bigcup_{j=1}^{6} B_j\right| + \sum_{i\in\{1,\ldots,6\}, j\in\{1,\ldots,6\}} \left|A_i \bigcap B_j\right|}$$

Per part (I)(a), the first two terms approximately equal $5.616\cdot 10^16$. Clearly

$$\bigcup_{i\in\{1,\ldots,6\}, j\in\{1,\ldots,6\}} \left(A_i \bigcap B_j\right) \subset \bigcup_{j=1}^{6} B_j$$

which implies

$$-\left|\bigcup_{j=1}^{6} B_j\right| + \left|A_i \bigcap_{i\in\{1,\ldots,6\}, j\in\{1,\ldots,6\}} B_j\right| < 0$$

so the answer to this part will be less than $5.616\cdot 10^16$, which makes sense.

Calculating $\left|\bigcup_{j=1}^{6} B_j\right|$ is not too difficult using Inclusion-Exclusion:

$$\left|\bigcup_{j=1}^{6} B_j\right| = \sum_{j=1}^{6}|B_j| - \sum_{1\le j<k\le 6}|B_j\cap B_k| + \sum_{1\le j<k<\ell\le 6}|B_j\cap B_k\cap B_\ell| - \ldots$$

$$+(-1)^{6-1}|B_1\cap B_2\cap B_3\cap\cdots\cap B_6|$$

where since

$$B_6 \subset B_5 \subset B_4 \subset B_3 \subset B_2 \subset B_1$$

which implies e.g.

$$B_1 \cap B_2 \cap B_3 \cap \cdots \cap B_6 = B_6,$$

$$\sum_{1 \le j < k \le 6} |B_j \cap B_j| = 5|B_6| + 4|B_5| + 3|B_4| + 2|B_3| + |B_2|$$

we have

$$\left| \bigcup_{j=1}^{6} B_j \right| = |B_6| + |B_5| + |B_4| + |B_3| + |B_2| + |B_1| - (5|B_6| + 4|B_5| + 3|B_4| + 2|B_3| + |B_2|)$$

$$+(4|B_6| + 3|B_5| + 2|B_4| + |B_3|) - (3|B_6| + 2|B_5| + |B_4|) + \ldots - |B_6|$$

$$= |B_1| - |B_2| + |B_3| - |B_4| + |B_5| - |B_6|$$

$$= \binom{7}{1}\binom{20}{1} \cdot (7-1)^{20-1} - \binom{7}{2}\binom{20}{2} \cdot 2! \cdot (7-2)^{20-2} + \binom{7}{3}\binom{20}{3} \cdot 3! \cdot (7-3)^{20-3} - \binom{7}{4}\binom{20}{4} \cdot 4! \cdot (7-4)^{20-4}$$

$$+ \binom{7}{5}\binom{20}{5} \cdot 5! \cdot (7-5)^{20-5} - \binom{7}{6}\binom{20}{6} \cdot 6!(7-6)^{20-6}$$

$$\boxed{\approx 5.846 \cdot 10^{16}}$$

However, calculating

$$\sum_{i \in \{1,\ldots,6\}, j \in \{1,\ldots,6\}} \left| A_i \bigcap B_j \right|$$

is very difficult because, for example, $B_2 \cap A_3$ is nonempty but $B_2 \not\subset A_3$ and $A_3 \not\subset B_2$.

(c) $\boxed{7^{20} \approx 7.979 \cdot 10^{16}}$

(II) (a) By Proposition 10, there are $\binom{19}{6} = \boxed{27,132}$ ways to do this.

(b) Similar to Problem 1 part (II)(c), if the vector $(x_1, x_2, \ldots, x_7)$ represents the number of gifts given to each child, we would like a solution such that

$$x_1 + x_2 + \ldots + x_7 = 20, x_i \ge 2 \ \forall i$$

By Proposition 12, the number of possible allocations under these conditions, is $\binom{20+7\cdot(1-2)-1}{7-1} = \binom{12}{6} = \boxed{924}$.

(c) By Proposition 11, the number of nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1 + x_2 + \ldots + x_r = n$$

is equal to $\binom{n+r-1}{r-1}$. In distributing 20 identical gifts to 7 different children, we can imagine the vector $(x_1, x_2, \ldots, x_{10})$ represents the number of gifts given to each child (where x_i is a nonnegative integer for all i). So we have $n = 20$ and $r = 7$. Therefore the number of possible allocations is

$$\binom{20+7-1}{7-1} = \boxed{165,765,600}$$

(III) Generalization of 1(I): If there are g distinguishable gifts and $c \geq g$ children, the number of distinct allocations if each child can receive

(a) at most one gift is $\binom{c}{g}g!$.

(b) at most two gifts is

$$\sum_{i=c-g+1}^{g} \binom{c}{i} \cdot \binom{i}{g-i} \cdot \frac{g!}{(2!)^{g-i}}$$

(c) any number of gifts is c^g.

Generalization of 1(II): If there are g identical gifts and $c \geq g$ children, the number of distinct allocations if each child can receive

(a) at most one gift is $\binom{c}{g}$.

(b) at most two gifts is

$$\sum_{i=c-g+1}^{g} \binom{c}{i} \cdot \binom{i}{g-i}$$

(c) any number of gifts is $\binom{g+c-1}{c-1}$

Generalization of 2(I): If there are g distinguishable gifts and $c \leq g$ children, the number of distinct allocations if each child must receive

(a) at least one gift is

$$c^g - \sum_{i=1}^{c-1} (-1)^{i+1} \binom{c}{i} \cdot (c-i)^g$$

(b) at least two gifts is

$$c^g - \sum_{i=1}^{c-1} (-1)^{i+1} \binom{c}{i} \cdot (c-i)^g - \sum_{i=1}^{c-1} (-1)^{i+1} \binom{c}{i} \binom{g}{i} \cdot i! \cdot (c-i)^{g-i}$$

(c) any number of gifts is c^g

Generalization of 2(II): If there are g identical gifts and $c \leq g$ children, the number of distinct allocations if each child must receive

(a) at least one gift is

$$\binom{g-1}{c-1}$$

(b) at least two gifts is

$$\binom{g-c-1}{c-1}$$

(c) any number of gifts is

$$\binom{g+c-1}{c-1}$$

Homework 1 Problem 4. You have \$20K to invest, and have a choice of stocks, bonds, mutual funds, or a CD. Investments must be made in multiples of \$1K, and there are minimal amounts to be invested: \$2K in stocks, \$2K in bonds, \$3K in mutual funds, and \$4K in the CD. Count the number of choices in each situation: (a) You want to invest in all four, (b) you want to invest in at least three out of four.

Solution.

(a) If the vector $(x_S, x_B, x_{MF}, x_{CD})$ represents the amount of money (in thousands of dollars) invested in each instrument, we would like a solution such that

$$x_S + x_B + x_{MF} + x_{CD} = 20$$

where

$$x_S \geq 2, x_B \geq 2, x_{MF} \geq 3, x_{CD} \geq 4$$

In a way similar to the proof for Proposition 12, note that we can transform this problem in the following way:

$$x_S - 1 + x_B - 1 + x_{MF} - 2 + x_{CD} - 3 = 20 - (1 + 1 + 2 + 3)$$

where

$$x_S - 1 \geq 1, x_B - 1 \geq 1, x_{MF} - 2 \geq 1, x_{CD} - 3 \geq 1$$

Letting $y_S = x_S - 1, y_B = x_B - 1, y_{MF} = x_{MF} - 2, y_{CD} = x_{CD} - 3$, we have the equivalent system

$$y_S + y_B + y_{MF} + y_{CD} = 13, y \geq 1 \ \forall y$$

By Proposition 10, the number of distinct solutions to this equation, and therefore the number of possible allocations under these conditions, is $\binom{13-1}{4-1} = \boxed{220}$.

(b) Enumerate the $\binom{4}{3} = 4$ possibilities.

(i) **Invest in stocks, bonds, and mutual funds.**

$$x_S + x_B + x_{MF} = 202$$

where

$$x_S \geq 2, x_B \geq 2, x_{MF} \geq 3$$

Note that we can transform this problem in the following way:

$$x_S - 1 + x_B - 1 + x_{MF} - 2 = 20 - (1 + 1 + 2)$$

where

$$x_S - 1 \geq 1, x_B - 1 \geq 1, x_{MF} - 2 \geq 1$$

Letting $y_S = x_S - 1, y_B = x_B - 1, y_{MF} = x_{MF} - 2$, we have the equivalent system

$$y_S + y_B + y_{MF} = 16, y \geq 1 \ \forall y$$

Therefore the number of possible allocations under these conditions is $\binom{16-1}{3-1} = \boxed{105}$.

(ii) **Invest in stocks, bonds, and CDs.**

$$x_S + x_B + x_{CD} = 20$$

where

$$x_S \geq 2, x_B \geq 2, x_{CD} \geq 4$$

Note that we can transform this problem in the following way:

$$x_S - 1 + x_B - 1 + x_{CD} - 3 = 20 - (1 + 1 + 3)$$

where

$$x_S - 1 \geq 1, x_B - 1 \geq 1, x_{CD} - 3 \geq 1$$

Letting $y_S = x_S - 1, y_B = x_B - 1, y_{CD} = x_{CD} - 3$, we have the equivalent system

$$y_S + y_B + y_{CD} = 15, y \geq 1 \ \forall y$$

Therefore the number of possible allocations under these conditions is $\binom{15-1}{3-1} = \boxed{91}$.

(iii) **Invest in stocks, mutual funds, and CDs.**

$$x_S + x_{MF} + x_{CD} = 2$$

where

$$x_S \geq 2, x_{MF} \geq 3, x_{CD} \geq 4$$

Note that we can transform this problem in the following way:

$$x_S - 1 + x_{MF} - 2 + x_{CD} - 3 = 20 - (1 + 2 + 3)$$

where

$$x_S - 1 \geq 1, x_{MF} - 2 \geq 1, x_{CD} - 3 \geq 1$$

Letting $y_S = x_S - 1, y_{MF} = x_{MF} - 2, y_{CD} = x_{CD} - 3$, we have the equivalent system

$$y_S + y_{MF} + y_{CD} = 14, y \geq 1 \ \forall y$$

Therefore the number of possible allocations under these conditions is $\binom{14-1}{3-1} = \boxed{78}$.

(iv) **Invest in bonds, mutual funds, and CDs.**

$$x_B + x_{MF} + x_{CD} = 2$$

where

$$x_B \geq 2, x_{MF} \geq 3, x_{CD} \geq 4$$

Note that we can transform this problem in the following way:

$$x_B - 1 + x_{MF} - 2 + x_{CD} - 3 = 20 - (1 + 2 + 3)$$

where

$$x_B - 1 \geq 1, x_{MF} - 2 \geq 1, x_{CD} - 3 \geq 1$$

Letting $y_B = x_B - 1, y_{MF} = x_{MF} - 2, y_{CD} = x_{CD} - 3$, we have the equivalent system

$$y_B + y_{MF} + y_{CD} = 14, y \geq 1 \ \forall y$$

therefore the number of possible allocations under these conditions is $\binom{14-1}{3-1} = \boxed{78}$.

(v) **Invest in all four:** per part 4(a), there are $\boxed{220}$ ways to do this.

Note that all of these possibilities are mutually exclusive. Therefore the total number is

$$\binom{16-1}{3-1} + \binom{15-1}{3-1} + \binom{14-1}{3-1} + \binom{14-1}{3-1} + \binom{13-1}{4-1} = 105 + 91 + 78 + 78 + 220 = \boxed{572}$$

1.3 To Know for Math 505A Midterm 2

1.3.1 Definitions

Definition 1.20. A random variable X is **continuous** if its distribution function $F(x) = \Pr(X \leq x)$ can be written as

$$F(x) = \int_{-\infty}^{x} f(u)du$$

for some integrable $f : \mathbb{R} \to [0, \infty)$.

Definition 1.21. The function f is called the **(probability) density function** of the continuous random variable X.

Proposition 34. If X has pdf $f_X(x)$, then for $\mu \in \mathbb{R}$, $\sigma > 0$,

$$h(x) = \frac{1}{\sigma} f_X\left(\frac{x - \mu}{\sigma}\right)$$

is a pdf. In this setting μ is sometimes called a "location parameter" and σ is called a "scale parameter."

Definition 1.22. The **joint distribution function** of X and Y is the function $F : \mathbb{R}^2 \to [0, 1]$ given by

$$F(x, y) = \Pr(X \leq x \cap Y \leq y)$$

Definition 1.23. The random variables X and Y are **jointly continuous** with **joint (probability) density function** $f : \mathbb{R}^2 \to [0, \infty)$ if

$$F(x, y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f(u, v)dudv \text{ for each } x, y \in \mathbb{R}$$

Definition 1.24. Two continuous random variables are **independent** if and only if $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events for all $x, y \in \mathbb{R}$.

Ways to show independence:

- Use Definition 1.24: show that $\Pr(X \leq x \cap Y \leq y) = \Pr(X \leq x)\Pr(Y \leq y)$ for all $x, y \in \mathbb{R}$.
-

  Theorem 35. The random variables X and Y are independent if and only if $F(x, y) = F_X(x)F_Y(x)$ for all $x, y \in \mathbb{R}$.
-

  Proposition 36. For continuous random variables, the previous condition is equivalent to requiring $f(x, y) = f_X(x)f_Y(y)$.
-

  Theorem 37. If two variables are bivariate normal, they are independent if and only if their covariance

  $$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)dxdy$$

  is equal to 0.

  Theorem 38. (Change of variables.) Let X_1, Y_1 be random variables with joint PDF f_{X_1,Y_1}. Let X_2, Y_2 be random variables with joint PDF f_{X_2,Y_2}. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ and let $S : \mathbb{R}^2 \to \mathbb{R}^2$ so that $ST(x, y) = (x, y)$ and $TS(x, y) = (x, y)$ for every $(x, y) \in \mathbb{R}^2$. Let $J(x, y)$ denote the determinant of the Jacobian of S at (x, y). Then

  $$f_{X_2,Y_2}(x, y) = f_{X_1,Y_1}(S(x, y)) \left|J(x, y)\right|.$$

  Proof. If the transformation from X_1, Y_1 to X_2, Y_2 is given by $S(X_1, Y_1) = (X_2, Y_2)$, then the change of variables formula from calculus is as follows:

  $$\int\int_A f_{X_1,Y_1}(x, y)dxdy = \int\int_B f_{X_1,Y_1}\big(S(x, y)\big) \left|J(x, y)\right| dxdy$$

  where $A \subseteq \mathrm{domain}(f_{X_1,Y_1}(\cdot))$, B is the transformation of the region A under S, and $|J(x, y)|$ is the Jacobian of S at (x, y). It follows from the definition of joint pdfs that the integrand on the right is the joint pdf of (X_2, Y_2); that is,

  $$f_{X_2,Y_2}(x, y) = f_{X_1,Y_1}\big(S(x, y)\big) \left|J(x, y)\right|.$$

  □
- Characteristic functions:

  Theorem 39. X and Y are independent if and only if $\phi_{X,Y}(s, t) = \phi_X(s)\phi_Y(t)$.

Theorem 40. (Theorem 4.2.3, Grimmett and Stirzaker.) Let X and Y be random variables, and let $g, h : \mathbb{R} \to \mathbb{R}$. If X and Y are independent, then so are $g(X)$ and $h(Y)$.

Definition 1.25. The **correlation coefficient** between random variables X and Y is given by

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Theorem 41. The correlation coefficient satisfies $|\rho| \leq 1$.

Proof. Apply the Cauchy-Schwarz Inequality (Theorem **??**) to $X - \mathbb{E}(X)$ and $Y - \mathbb{E}(Y)$:

$$\mathrm{Cov}(X,Y)^2 = \left(\mathbb{E}\left[\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\right]^2\right)^2 \leq \mathbb{E}\left[\big(X-\mathbb{E}(X)\big)^2\right]\mathbb{E}\left[\big(Y-\mathbb{E}(Y)\big)^2\right] = \mathrm{Var}(X)\mathrm{Var}(Y)$$

$$\iff \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} \leq 1 \iff \rho^2 \leq 1 \iff |\rho| \leq 1$$

$\square$

Theorem 42. (Stein identity.) Let X be a standard Gaussian random variable, so that X has density $x \to e^{-x^2/2}/\sqrt{2\pi}$, $\forall\, x \in \mathbb{R}$. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function such that g and g' have polynomial volume growth. That is, $\exists\, a, b > 0$ such that $|g(x)|, |g'(x)| \leq a(1+|x|)^b$, $\forall\, x \in \mathbb{R}$. Then

$$\mathbb{E}Xg(X) = \mathbb{E}g'(X).$$

Proof. Examining the left side, we have

$$\mathbb{E}(Xg(X)) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} xg(x)e^{-x^2/2}dx$$

We will use integration by parts with $u = g(x) \implies du = g'(x)dx, \;\; dv = xe^{-x^2/2}dx \implies v = -e^{-x^2/2}$ to yield the result:

$$\mathbb{E}(Xg(X)) = -g(x)e^{-x^2/2} + \int_{-\infty}^{\infty} g'(x)e^{-x^2/2}dx = \mathbb{E}(g'(X)).$$

$\square$

Example. Using Theorem 42, recursively compute $\mathbb{E}X^k$ for any positive integer k. Alternatively, for any $t > 0$, show that $\mathbb{E}e^{tX} = e^{t^2/2}$, i.e. compute the **moment generating function** of X. Then, using $\frac{d^k}{dt^k}|_{t=0}\mathbb{E}e^{tX} = \mathbb{E}X^k$ and using the power series expansion of the exponential, compute $\mathbb{E}X^k$ directly from the identity $\mathbb{E}e^{tX} = e^{t^2/2}$.

Solution. We can use this result to recursively calculate $\mathbb{E}(X^k)$ for any positive integer k. Suppose we have $\mathbb{E}(X^{k-1})$. Letting $g(X) = X^{k-1} \implies g'(X) = (k-1)X^{k-2}$, we have

$$\mathbb{E}(X^k) = \mathbb{E}(X \cdot X^{k-1}) = \mathbb{E}(Xg(X)) = \mathbb{E}(g'(X)) \iff \boxed{\mathbb{E}(X^k) = (k-1)\mathbb{E}(X^{k-2})}.$$

Since $X \sim \mathcal{N}(0,1)$, we have $\mathbb{E}(X) = 0, \quad \mathbb{E}(X^2) = \text{Var}(X) + \mathbb{E}(X)^2 = 1 + 0 = 1$. Therefore we have

$$\mathbb{E}(X^k) = \begin{cases} \prod_{i=1}^{(k-1)/2} (k - (2i-1))\mathbb{E}(X) & k \text{ is odd} \\ \prod_{i=1}^{k/2} (k - (2i-1))\mathbb{E}(X^2) & k \text{ is even} \end{cases}$$

$$\boxed{\mathbb{E}(X^k) = \begin{cases} 0 & k \text{ is odd} \\ \prod_{i=1}^{k/2} (k - (2i-1)) & k \text{ is even} \end{cases}}$$

1.3.2 Probability-Generating Functions

Definition 1.26.

$$G_X(s) = \mathbb{E}(s^X)$$

Theorem 43. Some useful properties:

(a) $\mathbb{E}(X) = G_X'(1)$, $\mathbb{E}[X(X-1)\cdots(X-k+1)] = G^{(k)}(1)$

(b) If X and Y are independent then $G_{X+Y}(s) = G_X(s)G_Y(s)$.

1.3.3 Moment-Generating Functions

Definition 1.27.

$$M_X(t) = \mathbb{E}(e^{tX})$$

Theorem 44. Some useful properties:

(a) $\mathbb{E}(X) = M_X'(0)$, $\mathbb{E}(X^k) = M^{(k)}(0)$

(b) If X and Y are independent then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

1.3.4 Characteristic Functions

Definition 1.28.

$$\phi_X(t) = \mathbb{E}(e^{itX})$$

Proposition 45. Necessary and sufficient conditions for a function to be a characteristic function:

(a) $\phi_X(0) = 1$

(b) $|\phi(t)| \leq 1 \ \forall \ t$

(c) ϕ is uniformly continuous on $\mathbb{R}$

(d) ϕ is positive semidefinite; that is,

$$\sum_{i,j} \phi(t_j - t_k) z_j \overline{z}_k \geq 0 \text{ for all real } t_1, t_2, \ldots, t_n \text{ and complex } z_1, z_2, \ldots, z_n$$

Or, equivalently, or every set of real numbers $t_1, t_2, \ldots, t_n$, the matrix $\phi(t_i - t_j), i, j \in \{1, 2, \ldots, n\}$ is Hermitian and nonnegative definite.

Remark. Relationship between characteristic functions and probability and moment generating functions:

$$\phi_X(t) = M_X(it) = G_X(e^{it})$$

Theorem 46. Some useful properties:

(a) $X \perp\!\!\!\perp Y \implies \phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$

(b) $Y = aX + b \implies \phi_Y(t) = e^{itb}\phi_X(at)$

(c) $\phi_X^{(k)}(0) = i^k \mathbb{E}(X^k)$

(d) $\phi_{X,Y}(s,t) = \mathbb{E}(e^{isX}e^{itY})$

(e) $X \perp\!\!\!\perp Y \iff \phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t)$

Theorem 47. Other facts from notes on course website

(a) If $\phi(t)$ is even, $\phi(0) = 1$, ϕ is convex for $t > 0$, and $\lim_{t\to\infty} \phi(t) = 0$, then ϕ is a characteristic function of an absolutely continuous random variable.

(b) If ϕ is a characteristic function and $\phi(t) = 1 + o(t^2), t \to 0$, then $\phi(t) = 1$ for all t. The random variable with such a characteristic function must have zero mean and zero variance. In particular, if $r > 2$, then $\exp(-|t|^r)$ is not a characteristic function.

(c) If $\phi(t) = e^{p(t)}$ is a characteristic function and $p = p(t)$ is a polynomial, then the degree of p is at most 2. For example, $e^{t^2 - t^4}$ is not a characteristic function.

(d) If ξ is absolutely continuous, then $\lim_{|t|\to\infty} |\phi_\xi(t)| = 0$ (Riemann-Lebesgue).

(e) If $\int_{-\infty}^{\infty} |\phi_\xi(t)|\, dt < \infty$, then ξ is absolutely continuous with pdf

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\phi(t)dt$$

1.3.5 Continuous Random Variable Distributions

Uniform: $\mathrm{U}(a, b)$

- Probability density function:

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

- Cumulative distribution function:

$$F(x) = \Pr(X \leq x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b \end{cases}$$

- Probability-generating function:
- Moment-generating function:

$$M_X(t) = \frac{1}{(b-a)t}\big[\exp(bt) - \exp(at)\big]$$

Proof.

$$M_X(t) = \mathbb{E}(\exp(tX)) = \int_a^b \frac{1}{b-a} \cdot \exp(tx)dx = \frac{1}{b-a}\left[\frac{1}{t}\exp(tx)\right]_a^b = \frac{1}{(b-a)t}\big[\exp(bt) - \exp(at)\big]$$

□

- Characteristic function:

$$\frac{2}{(b-a)t}\sin\left(\frac{1}{2}(b-a)t\right)\exp\left(i(a+b)\frac{t}{2}\right)$$

- Expectation: $\mathbb{E}(X) = (b-a)/2$
- Variance: $\text{Var}(X) = (b-a)^2/12$

Proposition 48. If $X \sim U(0,1)$, then $Y = -\log(X) \sim \text{Exponential}(1)$.

Proof.

$$\Pr(Y \leq y) = \Pr(-\log(X) \leq y) = \Pr(\log(X) \geq -y) = \Pr(X \geq e^{-y}) = \int_{\exp(-y)}^{\infty} dt = \int_{\exp(-y)}^{1} dt$$

Substituting $t = e^{-u}$ (so that we have $u = -\log(t)$, $dt = -e^{-u}du$, we have

$$\Pr(Y \leq y) = -\int_y^0 e^{-u}du = \left[e^{-u}\right]_y^0 = 1 - e^{-y}$$

which is the cdf for an exponential distribution with mean 1. □

Normal: $\mathcal{N}(\mu, \sigma^2)$

- Probability density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$

- Probability-generating function:
- Moment-generating function:
- Characteristic function: $\phi(t) = \exp(i\mu t - (1/2)\sigma^2 t^2)$. Standard normal: $\phi(t) = \exp((-1/2)t^2)$.
- Expectation: $\mathbb{E}(X) = \mu$
- Variance: $\text{Var}(X) = \sigma^2$

Gamma: $\Gamma(\alpha, \beta)$

- Probability density function:

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} = \frac{1}{\Gamma(\alpha, \beta)} x^{\alpha-1} e^{-x/\beta}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = \alpha\beta$
- Variance: $\text{Var}(X) = \alpha\beta^2$

Proposition 49. Let $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \ \sim \text{Gamma}(\alpha_2, \beta)$. Then $X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$.

Proposition 50. Let $X \sim \text{Gamma}(\alpha, \beta)$. Then as $\beta \to \infty$, $X \xrightarrow{d} \mathcal{N}(\alpha\beta, \alpha\beta^2)$.

Proof. See `http://www.math.wm.edu/~leemis/chart/UDR/PDFs/GammaNormal1.pdf`. □

χ^2_n: special case of a gamma distribution: $\Gamma(n/2, 2)$. Also the sum of n independent standard normally distributed variables.

- Probability density function:

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} = \frac{1}{\Gamma(n/2, 2)} x^{n/2-1} e^{-x/2}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = n/2 \cdot 2 = n$
- Variance: $\text{Var}(X) = n/2 \cdot 2^2 = 2n$

Exponential: (special case of a gamma distribution: $\Gamma(1, \beta)$. Also a special case of a Weibull distribution with $\beta = 1$.)

- Probability density function: $f(x) = \frac{1}{\beta}\exp(-x/\beta) = \lambda e^{-\lambda x}$
- Cumulative distribution function: $F(x) = \Pr(X \leq x) = 1 - e^{-\lambda x}$
- Probability-generating function:
- Moment-generating function: $\frac{\lambda}{\lambda - t}$
- Characteristic function:
- Expectation: $\mathbb{E}(X) = \beta = \lambda^{-1}$
- Variance: $\mathrm{Var}(X) = \beta^2 = \lambda^{-2}$

Remark. Recall Proposition 48: If $X \sim U(0,1)$, then $Y = -\log(X) \sim \text{Exponential}(1)$.

Cauchy:

- Probability density function:

$$f(x) = \frac{1}{\pi(1+x^2)} \text{ (standard Cauchy) }, f(x) = \frac{1}{\pi\sigma(1+(x-\mu)^2/\sigma^2)} \text{ (general)}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: does not exist
- Variance: does not exist (Cauchy distribution has no moments.)

Beta: Recall:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$\implies \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha+\beta}$$

- Probability density function: $f(x) =$
- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:

- Characteristic function:
- Expectation: $\mathbb{E}(X) =$
- Variance: $\text{Var}(X) =$

t_n:

- Probability density function:

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \cdot \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = 0$
- Variance: $\text{Var}(X) = n/(n-2)$

Weibull:

- Probability density function: $f(x) = \alpha\beta x^{\beta-1} \exp(-\alpha x^{\beta})$
- Cumulative distribution function: $F(x) = \Pr(X \leq x) = 1 - \exp(-\alpha x^{\beta})$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) =$
- Variance: $\text{Var}(X) =$

Pareto: (parameters α, x_m)

- Probability density function: $f(x) = \alpha x_m^{\alpha}/x^{\alpha+1}$ for $x \geq 1, 0$ otherwise.
- Cumulative distribution function: $F(x) = 1 - (x_m/x)^{\alpha}$ for $x \geq 1, 0$ otherwise.
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = \alpha x_m/(\alpha - 1)$ for $\alpha > 1, \infty$ otherwise.
- Variance: $\text{Var}(X) = \frac{x_m^2 \alpha}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$, ∞ otherwise.

1.3.6 Multivariate Gaussian (Normal) Distributions

Definition 1.29. From http://pluto.huji.ac.il/~pchiga/teaching/MathStat/SIAnotes2013.pdf (definition 2b6): A random vector $X = (X_1, X_2)$ is Gaussian with mean $\mu = (\mu_1, \mu_2)$ and the covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

if it has a joint pdf of the form

$$f_X(x) = \frac{1}{2\pi\sigma_2\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2)} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_1} + \frac{(x_2-\mu_2)^2}{\sigma_2^2)} \right)\right]$$

for $x \in \mathbb{R}^2$.

Proposition 51. From http://pluto.huji.ac.il/~pchiga/teaching/MathStat/SIAnotes2013.pdf (Proposition 3c1): Let X be a Gaussian random variable in $\mathbb{R}^2$ such that

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$$

Then $f_{X_1|X_2}(x_1; x_2)$ is Gaussian with the (conditional) mean

$$\mathbb{E}(X_1 \mid X_2 = x_2) = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

and the (conditional) variance

$$\text{Var}(X_1 \mid X_2 = x_2) = \sigma_1^2(1-\rho^2)$$

That is, the conditional distribution of X_1 given X_2 is

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2), (1-\rho^2)\sigma_1^2 \right)$$

Remark. Note that this matches the OLS coefficients in the univariate case. In other words, the univariate OLS formula can be derived using only this fact.

Recall Theorem 37: if two variables are bivariate normal, they are independent if and only if their covariance

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$$

equals 0.

1.4 Worked problems

1.4.1 Example Problems That Will Likely Appear on Midterm (and Final)

(1) Let X be uniform on $[1, 5]$, let Y be uniform on $[0, 1]$, and assume that X and Y are independent.

(a) Compute the probability density function of the product XY.

(b) **Only part included on midterm.** Compute the cumulative distribution function of the ratio X/Y.

(c) Compute the characteristic function of the sum $X + Y$.

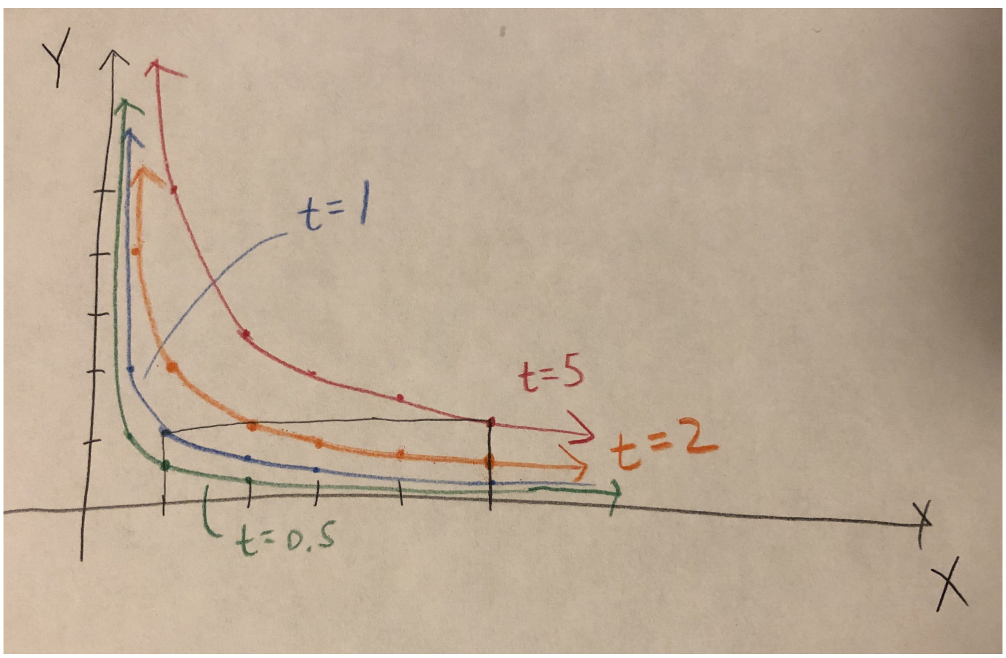
(d) Compute the moment-generating function of the random variable $X - \ln(Y)$.

Solution.

(a) We will find the cdf and then differentiate to yield the pdf. Observe that

$$F_{XY}(t) = \Pr(XY \leq t) = \Pr(Y \leq tX^{-1})$$

Plotting the density function of XY along with plots of $F_{XY}(t)$ as a function of X for various values of t, we have the following:



Now since both X and Y are distributed uniformly, for a given t, $\Pr(Y \leq tX^{-1})$ is the area under the curve and in the rectangle, weighted by 1/4 since the rectangle has total area 4 but total probability 1. It is clear that the four regimes we need to consider are (1) $t < 0$, (2) $0 \leq t < 1$, (3) $1 \leq t < 5$, and (4) $t \geq 5$.

(1) $t < 0$: The curve lies below the rectangle, so there is no area below the curve and in the rectangle. Therefore $\boxed{\Pr(Y \leq tX^{-1} \mid t < 0) = 0}$. (This is also clear since tX^{-1} would be a negative number and Y is nonnegative.)

(2) $0 \le t < 1$: Integrating the relevant area, we have

$$\Pr(Y \le tX^{-1} \mid 0 \le t < 1) = \frac{1}{4}\int_1^5 \frac{t}{x}dx = \frac{t}{4}\big[\log(x)\big]_1^5 = \boxed{\frac{t}{4}\log(5)}$$

(3) $1 \le t < 5$: In this case, the area is a rectangle of height 1 and width $t-1$ plus the area under the curve from t to 5.

$$\Pr(Y \le tX^{-1} \mid 1 \le t < 5) = \frac{1}{4}\left(1\cdot(t-1)+\int_t^5 \frac{t}{x}dx\right) = \frac{1}{4}\left(t-1+t\big[\log(x)\big]_t^5\right) = \boxed{\frac{1}{4}\big[t\big(1+\log(5/t)\big)-1\big]}$$

(4) $t \ge 5$: In this case, the entire rectangle lies below the curve. Therefore $\boxed{\Pr(Y \le tX^{-1} \mid t \ge 5) = 1}$.

So we have

$$F_{XY}(t) = \begin{cases} 0 & t < 0 \\ \frac{t}{4}\log(5) & 0 \le t < 1 \\ \frac{1}{4}\big[t\big(1+\log(5/t)\big)-1\big] & 1 \le t < 5 \\ 1 & t \ge 5 \end{cases}$$

Finally, differentiating yields

$$\boxed{f_{XY}(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{4}\log(5) & 0 \le t < 1 \\ \frac{1}{4}\log\left(\frac{5}{t}\right) & 1 \le t < 5 \\ 0 & t \ge 5 \end{cases}}$$

since

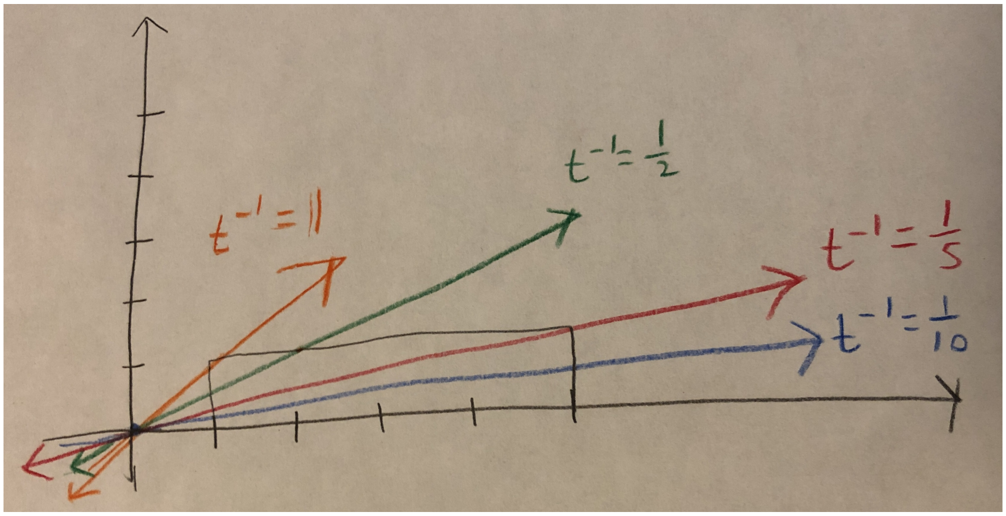
$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{4}\big[t\big(1+\log(5/t)\big)-1\big]\right) = \frac{1}{4}\left[1+\log(5/t)+t\left(\frac{1}{5/t}\cdot -5\cdot t^{-2}\right)\right] = \frac{1}{4}\left[1+\log(5/t)+t\left(t\cdot -1\cdot t^{-2}\right)\right]$$

$$= \frac{1}{4}\log\left(\frac{5}{t}\right)$$

(b) **Only part included on midterm.** We will proceed in a similar way as part (a). Observe that

$$F_{X/Y}(t) = \Pr\left(\frac{X}{Y} \le t\right) = \Pr(Y \ge X/t)$$

Plotting the density function of X/Y along with plots of $F_{X/Y}(t)$ as a function of X for various values of t, we have the following:



Now since both X and Y are distributed uniformly, for a given t, $\Pr(Y \geq X/t)$ is the area above the curve and in the rectangle, weighted by 1/4 since the rectangle has total area 4 but total probability 1. It is clear that the three regimes we need to consider are (1) $t^{-1} \geq 1 \iff t \leq 1$, (2) $1/5 \leq t^{-1} < 1 \iff 1 < t \leq 5$, and (3) $0 < t^{-1} < 1/5 \iff t > 5$.

(1) $t \leq 1$: The curve lies above the rectangle, so there is no area above the curve and in the rectangle. Therefore $\Pr(Y \geq X/t \mid t \leq 1) = 0$. (This is also clear since X/t would have to be greater than 1 and Y is less than or equal to 1.)

(2) $1 < t \leq 5$: The relevant area is the triangle above the green line in the rectangle. Note that it intersects the vertical line at $Y = 1/t$ and the horizontal line at $X = t$.

$$\Pr(Y \geq X/t \mid 1 < t \leq 5) = \frac{1}{4} \cdot \frac{1}{2}\left(1 - \frac{1}{t}\right)(t-1) = \frac{1}{8}\left(t - 1 - 1 + \frac{1}{t}\right) = \frac{1}{8}\left(t - 2 + \frac{1}{t}\right)$$

(3) $t > 5$: In this case, the area is a trapezoid above the blue line and in the rectangle. Note that the blue line intersects the left vertical line at $Y = 1/t$ and the right vertical line at $Y = 5/t$.

$$\Pr(Y \geq X/t \mid t > 5) = \frac{1}{4} \cdot \frac{1}{2} \cdot \left(1 - \frac{1}{t} + 1 - \frac{5}{t}\right) \cdot 4 = \frac{1}{2} \cdot \left(2 - \frac{6}{t}\right) = 1 - \frac{3}{t}$$

So we have

$$\boxed{F_{XY}(t) = \begin{cases} 0 & t \leq 1 \\ \frac{1}{8}\left(t - 2 + \frac{1}{t}\right) & 1 < t \leq 5 \\ 1 - \frac{3}{t} & t > 5 \end{cases}}$$

(c) The characteristic function for a uniform distribution on $[a, b]$ is

$$\frac{2}{(b-a)t} \sin\left(\frac{1}{2}(b-a)t\right) \exp\left(i(a+b)\frac{t}{2}\right).$$

Using the fact that $X \perp\!\!\!\perp Y \implies \phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$, we have

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \frac{2}{(5-1)t} \sin\left(\frac{1}{2}(5-1)t\right) \exp\left(i(5+1)\frac{t}{2}\right) \cdot \frac{2}{t} \sin\left(\frac{1}{2}t\right) \exp\left(i\frac{t}{2}\right)$$

$$= \frac{1}{2t}\sin(2t)\exp(3it)\cdot\frac{2}{t}\sin\left(\frac{1}{2}t\right)\exp\left(i\frac{t}{2}\right) = \boxed{\frac{1}{t^2}\exp\left(\frac{7}{2}it\right)\cdot\sin(2t)\sin\left(\frac{1}{2}t\right)}$$

(d) The moment-generating function for a uniform distribution on $[a,b]$ is

$$M_X(t) = \mathbb{E}(\exp(tX)) = \int_a^b \frac{1}{b-a}\cdot\exp(tx)dx = \frac{1}{b-a}\left[\frac{1}{t}\exp(tx)\right]_a^b = \frac{1}{(b-a)t}\big[\exp(bt)-\exp(at)\big]$$

Therefore the moment-generating function for X is $t^{-1}\big[\exp(t)-1\big]$. Note that

$$\Pr(Y\le y) = \Pr(-\log(X)\le y) = \Pr(\log(X)\ge -y) = \Pr(X\ge e^{-y}) = \int_{\exp(-y)}^{\infty} dt = \int_{\exp(-y)}^{1} dt$$

Substituting $t = e^{-u}$ (so that we have $u = -\log(t)$, $dt = -e^{-u}du$, we have

$$\Pr(Y\le y) = -\int_y^0 e^{-u}du = \left[e^{-u}\right]_y^0 = 1-e^{-y}$$

which is the cdf for an exponential distribution with mean 1. Therefore $Y = -\log(X)\sim \text{Exponential}(1)$, so

$$M_Y(t) = \frac{1}{1-t}$$

Using the fact that if X and Y are independent then $M_{X+Y}(t) = M_X(t)M_Y(t)$, we have

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \frac{\exp(t)-1}{t}\cdot\frac{1}{1-t} = \boxed{\frac{\exp(t)-1}{t-t^2}}$$

(2) **Fall 2016 Problem 2.** Let X and Y be i.i.d. exponential with mean 1. Show that for every $t>0$ the events $\{\omega : \min\{X,Y\}>t\}$ and $\{\omega : X<Y\}$ are independent.

Solution. Note that $\min\{X,Y\}>t \iff X>t\cap Y>t$.

- $\Pr(\min\{X,Y\}>t) = \Pr(X>t\cap Y>t) = \Pr(X>t)\Pr(Y>t)$

$$= \int_t^\infty e^{-x}dx\int_t^\infty e^{-y}dy = -e^{-x}\big|_t^\infty - e^{-y}\big|_t^\infty = \boxed{e^{-2t}}$$

- $\Pr(X<Y)$: Note that in Figure 2, the region that satisfies this condition is region G_1 plus G_2 (note that X and Y are nonnegative). Therefore we can find this probability by integrating the joint pdf over that region.

$$\Pr(X<Y) = \iint_{\{G_1+G_2\}} f_{X,Y}(x,y)dxdy$$

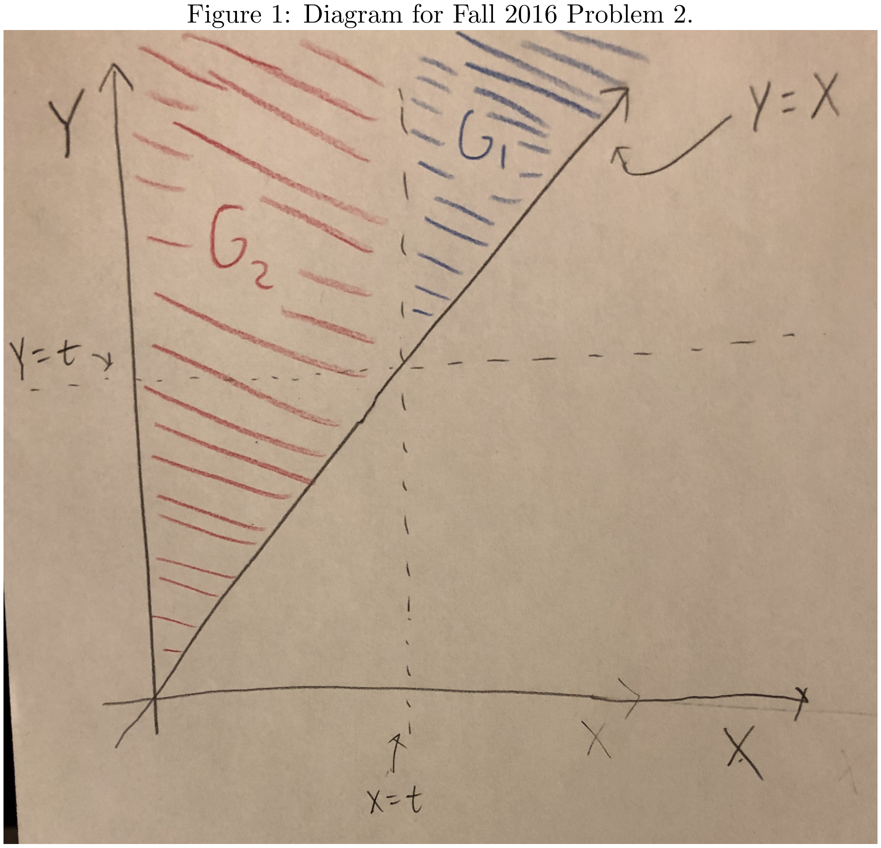
Note that the joint pdf is the probability of the marginal pdfs since X and Y are independent.

$$= \int_0^\infty\int_x^\infty e^{-x-y}dydx = \int_0^\infty e^{-x}\int_x^\infty e^{-y}dydx$$

$$\int_x^\infty e^{-y}dy = -e^{-y}\big|_x^\infty = e^{-x}$$

$$\implies \Pr(X<Y) = \int_0^\infty e^{-2x}dx = -\frac{1}{2}e^{-2x}\big|_0^\infty = \boxed{\frac{1}{2}}$$

Figure 1: Diagram for Fall 2016 Problem 2.



- $\Pr(X < Y \cap \min\{X, Y\} > t)$: Note that in Figure 2, the region that satisfies this condition is region G_1. Therefore we can find this probability by integrating the joint pdf over that region.

$$\Pr(X < Y \cap \min\{X,Y\} > t) = \iint\limits_{G_1} f_{X,Y}(x,y)dxdy = \int_t^\infty \int_t^y e^{-x-y}dxdy = \int_t^\infty e^{-y} \int_t^y e^{-x}dxdy$$

$$\int_t^y e^{-x}dx = -e^{-x}\big|_t^y = -e^{-y} + e^{-t}$$

$$\implies \Pr(X < Y \cap \min\{X,Y\} > t) = \int_t^\infty e^{-y}\big(-e^{-y} + e^{-t}\big)dy = \int_t^\infty \big(e^{-t-y} - e^{-2y}\big)dy$$

$$= \frac{1}{2}e^{-2y} - e^{-t}e^{-y}\big|_t^\infty = -\frac{1}{2}e^{-2t} - e^{-2t} = \boxed{\frac{1}{2}e^{-2t}}$$

Note that

$$\Pr(X < Y \cap \min\{X,Y\} > t) = \frac{1}{2} \cdot e^{-2t} = \Pr(X < Y)\Pr(\min\{X,Y\} > t)$$

Therefore the events $\{\omega : \min\{X, Y\} > t\}$ and $\{\omega : X < Y\}$ are independent for every $t > 0$.

(3) In a certain area, earthquakes happen at a frequency of one every four days. What is the probability that more than 100 earthquakes will occur in this area in one year (365 days)?

Solution. We can think of this as a Poisson process (see section **??**) with $\lambda = 1/4$. Then there are two ways to obtain the answer: we can either examine the number of earthquakes in a 365 day period $N(365)$ and find the probability that $N(365) > 100$, or we can examine the number of days until the 101st earthquake T_{101} and find the probability that $T_{101} < 365$.

(i) **Number of earthquakes in 365 days:** Let $N(t)$ be the number of earthquakes that occur in t days after the start of this process. By Theorem **??**, $N(t) \sim \text{Poisson}(t \cdot 1/4)$. Then

$$\Pr(N(t) > 100) = \sum_{j=101}^{\infty} \frac{(365 \cdot 1/4)^j \exp(-365 \cdot 1/4)}{j!}$$

To obtain an answer for this, we can use the normal approximation to a Poisson distribution (Proposition 27):

$$N(t) \sim \mathcal{N}(t/4, t/4) \implies \Pr(N(365) > 100) \approx \Pr\left(\mathcal{N}(0,1) > \frac{100.5 - 365/4}{\sqrt{365/4}}\right)$$

$$= \Pr\left(\mathcal{N}(0,1) > \frac{100.5 - 91.25}{\sqrt{91.25}}\right) \approx \Pr\left(\mathcal{N}(0,1) > \frac{9.25}{9.1}\right) \approx \boxed{0.1664}$$

(ii) **Number of days before 100th earthquake:** Let T_n be the number of days until the nth earthquake happens. By Corollary **??**, $T_n \sim \text{Gamma}(n, 4)$. Then

$$\Pr(T_{101} < 365) = \int_0^{365} \frac{1}{\Gamma(101,4)} x^{101-1}e^{-x/4}dx$$

To obtain an answer for this, we can use the normal approximation to a Gamma distribution (Proposition 50):

$$T_n \sim \mathcal{N}(404, 1616) \implies \Pr(T_{101} < 365) \approx \Pr\left(\mathcal{N}(0,1) < \frac{365 - 404}{\sqrt{1616}}\right) \approx \Pr\left(\mathcal{N}(0,1) < \frac{-40}{40}\right)$$

$$= \Pr\big(\mathcal{N}(0,1) < -1\big) \approx \boxed{0.1660}$$

1.4.2 More Problems From Homework

Homework 5 Problem 4.

Let $X_1, X_2, \ldots$ be i.i.d. having moment-generating functions $M_X = M_X(t), t \in (-\infty, \infty)$. Let N be an integer-valued random variable with moment-generating function $M_N = M_N(t), t \in (-\infty, \infty)$. Assume that N is independent of all X_k and define $S = \sum_{k=1}^{N} X_k$. Confirm that the random variable S has the moment-generating function $M_S = M_S(t)$ defined for all $t \in (-\infty, \infty)$ and

$$M_S(t) = M_N\big(M_X(t)\big)$$

Then use the result to derive the formulae

$$\mathbb{E}(S) = \mu_N \mu_X, \mathrm{Var}(S) = (\sigma_N^2 - \mu_N)\mu_X^2 + \mu_N \sigma_X^2$$

where $\mu_N = \mathbb{E}(N)$, $\mu_X = \mathbb{E}(X_1)$, $\sigma_N^2 = \mathrm{Var}(N)$, and $\sigma_X^2 = \mathrm{Var}(X_1)$. How will the above computations change if we use the characteristic function ϕ_X instead of the moment-generating function M_X?

Solution.

$$M_S(t) = \mathbb{E}(e^{tS}) = \mathbb{E}[\mathbb{E}(e^{tS} \mid N)] = \sum_{n=0}^{\infty} \mathbb{E}(e^{tS} \mid N = n) \Pr(N = n) = \sum_{n=0}^{\infty} \mathbb{E}(e^{t(X_1 + X_2 + \ldots + X_n)} \mid N = n) \Pr(N = n)$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{tX_1} e^{tX_2} \cdots e^{tX_n}) \Pr(N = n)$$

By independence of the X_i we have

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{tX_1}) \mathbb{E}(e^{tX_2}) \cdots \mathbb{E}(e^{tX_n}) \Pr(N = n)$$

which, since the X_i are i.i.d., can be written as

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{tX_1})^n \Pr(N = n) = \sum_{n=0}^{\infty} (M_X(t))^n \Pr(N = n)$$

But since $G_N(s) = \mathbb{E}(s^N) = \sum_{n=0}^{\infty} s^n \Pr(N = n)$, this can be written as

$$M_S(t) = G_N(M_X(t))$$

as desired. Note that

$$M'_S(t) = G'_N\big(M_X(t)\big)M'_X(t)$$

$$M''_S(t) = G''_N\big(M_X(t)\big)(M'_X(t))^2 + G'_N\big(M_X(t)\big)M''_X(t)$$

So we have

- $\mathbb{E}(S) = M'_S(0) = G'_N\big(M_X(0)\big)M'_X(0) = G'_N(1)\mathbb{E}(X_1) = \mathbb{E}(N)\mathbb{E}(X_1) = \mu_N\mu_X$
- $\text{Var}(S) = \mathbb{E}(S^2) - \mathbb{E}(S)^2 = M''_S(0) - (M'_S(0))^2$

$$= G''_N\big(M_X(0)\big)(M'_X(0))^2 + G'_N\big(M_X(0)\big)M''_X(0) - \mu_N^2\mu_X^2 = G''_N\big(1\big)\mathbb{E}(X_1)^2 + G'_N\big(1\big)\text{Var}(X_1) - \mu_N^2\mu_X^2$$

$$= \mathbb{E}[N(N-1)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2\mu_X^2 = \mathbb{E}[N^2 - N]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2\mu_X^2$$

$$= [\mathbb{E}(N^2) - \mathbb{E}(N)^2 + \mathbb{E}(N)^2 - \mathbb{E}(N)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2\mu_X^2 =$$

$$= [\text{Var}(N) + \mathbb{E}(N)^2 - \mathbb{E}(N)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2\mu_X^2 = (\sigma_N^2 + \mu_N^2 - \mu_N)\mu_X^2 + \mu_N\sigma_X^2 - \mu_N^2\mu_X^2$$

$$= \boxed{(\sigma_N^2 - \mu_N)\mu_X^2 + \mu_N\sigma_X^2}$$

To use the characteristic function ϕ_X instead of the moment generating function M_X, we would do the following:

$$\phi_S(t) = \mathbb{E}(e^{itS}) = \mathbb{E}[\mathbb{E}(e^{itS} \mid N)] = \sum_{n=0}^{\infty} \mathbb{E}(e^{itS} \mid N = n)\Pr(N = n) = \sum_{n=0}^{\infty} \mathbb{E}(e^{it(X_1+X_2+\ldots+X_n)} \mid N = n)\Pr(N = n)$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{itX_1}e^{tX_2}\cdots e^{itX_n})\Pr(N = n)$$

By independence of the X_i we have

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{itX_1})\mathbb{E}(e^{itX_2}) \cdots \mathbb{E}(e^{itX_n}) \Pr(N = n)$$

which, since the X_i are i.i.d., can be written as

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{itX_1})^n \Pr(N = n) = \sum_{n=0}^{\infty} (\phi_X(t))^n \Pr(N = n)$$

But since $G_N(s) = \mathbb{E}(s^N) = \sum_{n=0}^{\infty} s^n \Pr(N = n)$, this can be written as

$$\phi_S(t) = G_N(\phi_X(t))$$

Homework 5 Problem 7.

(a) Let $X_1, X_2, \ldots, X_n$ be independent with mean zero and finite third moment. Prove that

$$\mathbb{E}(X_1 + \ldots + X_n)^3 = \mathbb{E}X_1^3 + \ldots + \mathbb{E}X_n^3$$

Solution.

(a) Let $\mathbb{E}(\exp(it_iX_i) = \phi_{X_i}(t_i)$. Let $S_n = \sum_{i=1}^n X_i$. Then by independence the characteristic function for S_n is

$$\mathbb{E}(\exp(itS_n)) = \phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_i}(t)$$

Then

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n)^3 = \mathbb{E}(S_n^3) = \phi_{S_n}^{(3)}(0)$$

$$= \sum_{i=1}^n \phi_{X_i}^{(3)}(0) \cdot \left(\prod_{j \in \{1,\ldots,n\}, j \neq i} \phi_{X_j}(0) \right) + C \left[\sum_{i=1}^n \cdot \left(\sum_{j \in \{1,\ldots,n\}, j \neq i} \phi_{X_i}^{(2)}(0) \phi_{X_j}^{(1)}(0) \right) \cdot \left(\prod_{k \in \{1,\ldots,n\}, k \neq i,j} \phi_{X_k}(0) \right) \right]$$

where C is some coefficient resulting from the multinomial expansion of S_n after repeated differentiation product rules. But because $\mathbb{E}(X_i) = 0$, $\phi_{X_i}^{(1)}(0) = 0 \ \forall i$, so the second term goes to 0. Therefore we have

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n)^3 = \sum_{i=1}^n \phi_{X_i}^{(3)}(0) \cdot \left(\prod_{j \in \{1,\ldots,n\}, j \neq i} \phi_{X_j}(0) \right) = \sum_{i=1}^n \mathbb{E}(X_i^3) \cdot 1^{n-1} = \sum_{i=1}^n \mathbb{E}(X_i^3)$$

as desired.

Homework 6 Problem 10.

(a) For $p \in (0, 1)$, let $x(p)$ be the smallest number of people so that there is a better than $100 \cdot p\%$ chance to have at least two born on the same day. Find an approximate expression for $x(p)$, and sketch the graph of the function $x = x(p)$.

(b) Repeat part (a) when you want at least three people to share a birthday.

Solution.

(a) Let $f(x)$ be the probability of no matches in birthdays in a group of x people; that is,

$$f(x) = \frac{365 \cdot 364 \cdot 363 \cdots (365 - x + 1)}{365^x} = \frac{1}{365^x} \cdot \frac{365!}{(365 - x)!} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{x-1}{365}\right)$$

Using the first order Taylor approximation $\exp(-k/x) \approx 1 - k/x$, we have

$$f(x) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{x-1}{365}\right) \approx \exp(-1/365)\exp(-2/365) \cdots \exp(-(x-1)/365)$$

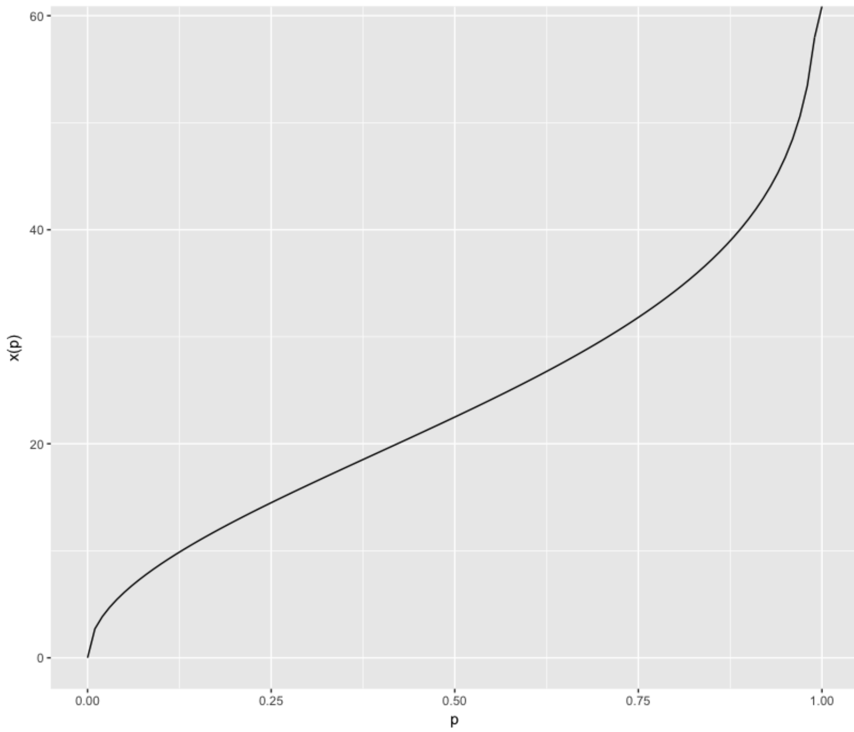
$$= e^{-(x^2 - x)/(2 \cdot 365)}$$

We want the probability of a match to be at least p; that is, $f(x) \le 1 - p$. Setting this equal to $q = 1 - p$, we have

$$e^{-(x^2-x)/(2\cdot 365)} = q \iff -\frac{x^2 - x}{730} = \log(q) \iff x^2 - x + 730\log(q) = 0$$

$$\implies x = 0.5 + \sqrt{1/4 + 730\log(1/q)} \approx \boxed{\sqrt{2 \cdot 365 \log(1/q)}}$$

where we discard the negative root because we have to have a nonnegative number of people, and we don't worry about the decimals since this is an approximation and we have to round up to the nearest whole person anyway.



(b) For a group of three people, the Poisson approximation (see Section 1.1.8) is more convenient. The number of groups of 3 people in a room of x people is $\binom{x}{3}$. For a group of three people, the probability that all three have the same birthday is $1 \cdot 1/365 \cdot 1/365 = 365^{-2}$. Therefore we can think of the number of matches of three people as distributed Poisson with expectation $\binom{x}{3} \cdot 365^{-2}$. Then we have the probability of at least one "success" (triplet with three matched birthdays) is

$$1 - \frac{\exp(-\lambda)\lambda^0}{0!} = 1 - \exp\left(-\binom{x}{3} \cdot 365^{-2}\right)$$

We set this equal to p and solve:

$$p = 1 - \exp\left(-\binom{x}{3} \cdot 365^{-2}\right) \iff -\binom{x}{3} \cdot 365^{-2} = \log(1-p) \iff \frac{x!}{(x-3)!3!} = 365^2 \cdot \log\left(\frac{1}{1-p}\right)$$

$$\iff x(x-1)(x-2) = 6 \cdot 365^2 \cdot \log\left(\frac{1}{1-p}\right) \iff (x^2-x)(x-2) = x^3 - 3x^2 + 2x = 6 \cdot 365^2 \cdot \log\left(\frac{1}{1-p}\right)$$

This has a unique real solution, but it is hard to find.

Question:

Let X, Y, Z be independent uniform on (0, 1). Compute the cdfs of XY, X/Y, and XY/Z.

Solution.

Using the information from part (a), and the fact that $f_X(x) = 1$ (for $x \in [0, 1]$) and likewise for $f_Y(y)$:

- XY:

$$F_{XY}(z) = \int_0^\infty f_X(x) \int_{-\infty}^{z/x} f_Y(y) dy dx - \int_{-\infty}^0 f_X(x) \int_\infty^{z/x} f_Y(y) dy dx$$

$$= \int_0^1 \left[(z/x)\mathbf{1}_{\{0<z/x\le 1\}} + \mathbf{1}_{\{z/x>1\}}\right] dx = \int_0^1 \left[(z/x)\mathbf{1}_{\{z\le x\}} + \mathbf{1}_{\{z>x\}}\right] dx = \int_0^z dx + \int_z^1 (z/x) dx$$

$$= z + z\log(x)\Big|_z^1 = z + z\log(1) - z\log(z) = z(1-\log(z))$$

$$\implies \boxed{F_{XY}(z) = \begin{cases} 0 & z \le 0 \\ z(1-\log(z)) & 0 < z \le 1 \\ 1 & z > 1 \end{cases}}$$

- X/Y:

$$F_{X/Y}(z) = \int_0^\infty f_Y(y) \int_{-\infty}^{zy} f_X(x) dx dy - \int_{-\infty}^0 f_Y(y) \int_\infty^{zy} f_X(x) dx dy$$

$$= \int_0^1 \left[zy\mathbf{1}_{\{0<zy\le 1\}} + \mathbf{1}_{\{zy>1\}}\right] dy = \int_0^1 \left[zy\mathbf{1}_{\{y>0\cap y\le 1/z\}} + \mathbf{1}_{\{y>1/z\}}\right] dy = \int_0^{1/z} zy \cdot dy + \int_{1/z}^1 dy$$

$$= \frac{zy^2}{2}\Bigg|_0^{1/z} + (1-1/z) = \frac{z}{2z^2} + 1 - \frac{2}{2z} = 1 - \frac{1}{2z}$$

$$\implies F_{XY}(z) = \begin{cases} 0 & z \le 0 \\ 1 - \frac{1}{2z} & 0 < 1/z \le 1 \end{cases}$$

$$= \boxed{\begin{cases} 0 & z \le 0 \\ z/2 & 0 < z \le 1 \\ 1 - \frac{1}{2z} & z > 1 \end{cases}}$$

- XY/Z: Consider this the cdf of the quotient of $W = XY$ and Z.

$$F_U(u) = \int_0^\infty f_Z(z) \int_{-\infty}^{uz} f_W(w) dw dz - \int_{-\infty}^0 f_Z(z) \int_\infty^{uz} f_W(w) dw dz$$

$$= \int_0^1 \int_0^{uz} -\log(w)\mathbf{1}_{\{0<uz\le 1\}} dw dz = \int_0^1 -\big[w\log(w) - w\big]_0^{uz} \mathbf{1}_{\{0<z\le 1/u\}} dz$$

$$= \int_0^{1/u} uz\big[1-\log(uz)\big] dz = \frac{u}{4} z^2\big(3 - 2\log(uz)\big)\Bigg|_0^{1/u} = \frac{u}{4u^2}\big(3-2\log(1)\big) - 0 = \frac{3}{4u}$$

$$\implies \boxed{F_{XY/Z}(u) = \begin{cases} 0 & u \le 0 \\ \frac{3}{4u} & 0 < u \le 3/4 \\ 1 & u > 3/4 \end{cases}}$$