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 5

1.1.4 Methods for Calculating Quantities 6

1.1.5 Discrete Random Variable Distributions 7

1.1.6 Indicator Method 9

1.1.7 Linear transformations of random variables 9

1.1.8 Poisson Paradigm (Poisson approximation for indicator method) 9

1.1.9 Asymptotic Distributions 10

1.2 Worked problems 10

1.2.1 Example Problems That Will Likely Appear on Midterm 10

1.2.2 Problems we did in class that professor mentioned 15

1.2.3 Problems we did on homework 17

1.3 To Know for Math 505A Midterm 2 30

1.3.1 Definitions 30

1.3.2 Probability-Generating Functions 32

1.3.3 Moment-Generating Functions 32

1.3.4 Characteristic Functions 32

1.3.5 Continuous Random Variable Distributions 33

1.3.6 Multivariate Gaussian (Normal) Distributions 37

1.4 Worked problems 38

1.4.1 Example Problems That Will Likely Appear on Midterm 38

1.4.2 More Problems From Homework 39

Last updated November 27, 2018

1 Probability

These are my notes from taking Math 505A at USC and the textbook *Probability and Random Processes* (Grimmet and Stirkazer) 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. 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 1.7. 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 $\operatorname{Cov}(X,Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ 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 1.8. 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 1.9. The **conditional probability mass function** of Y given $X = x$, written $f_{Y|X}(\cdot \mid x)$, is defined by

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

Proposition 4. Iterated expectations:

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

Definition 1.10. Conditional Variance: $\mathrm{Var}(X \mid Y) = \mathbb{E}\big[(X - \mathbb{E}(X \mid Y))^2 \mid Y\big]$

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)

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

Theorem 6. 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 7. Variance-Covariance Expansion. Let $X_1, \ldots, X_n$ be random variables. If $\mathbb{E}\left|X_k\right|^2 < \infty$, then

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

Proposition 8. (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 9. (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 $(x_1, x_2, \ldots, x_r), x_i > 0 \; \forall \; i$ 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

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

By Proposition 8, the number of distinct solutions to this equation is $\binom{n+r-1}{r-1}$.

□

1.1.4 Methods for Calculating Quantities

- Expectation

  –

  Definition 1.11. $\mathbb{E}(X) = \sum_x x \Pr(X = x)$

  –

  Theorem 10. (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 11. 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 12. Expectation is a linear operator: $\mathbb{E}\big(\sum_i X_i\big) = \sum_i \mathbb{E}(X_i)$

- Variance

  –

  Definition 1.12. $\text{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$

  –

  Proposition 13. (Useful reformulation:) $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

  –

  Theorem 14. (Some useful results):

  (a) $\text{Var}(aX) = a^2\text{Var}(X)$

  (b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

  (c) $\text{Var}(aX \pm bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) \pm 2ab\text{Cov}(X, Y)$

  –

  Theorem 15. Law of Total variance: $\text{Var}(X) = \text{Var}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}\big(\text{Var}(X \mid Y)\big)$

- Covariance

  –

  Definition 1.13. $\text{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big]$

  –

  Proposition 16. (Useful reformulation): $\text{Cov}(X) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

  –

  Definition 1.14. 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 17. 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: Binomial(n, p) (sum of n Bernoulli random variables)

- Mass function: $\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- Distribution: $\Pr(X \le k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1 - p)^{n-i}$
- 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: $\text{Var}(X) = \lambda$

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

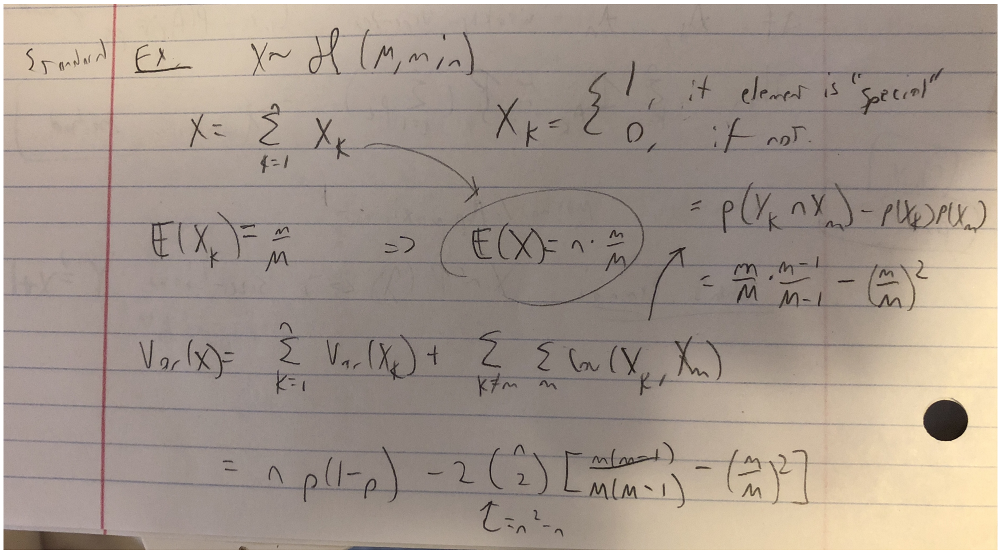
- Mass function: $\Pr(X = k) = p(1-p)^{k-1}$
- Distribution: $\Pr(X \leq k) = \sum_{i=1}^{k} p(1-p)^{k-1}$
- Expectation: $\mathbb{E}(X) = 1/p$
- Variance: $\text{Var}(X) = (1-p)/p^2$

Negative binomial: $\text{NB}(r, p)$: 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: $\mathbb{E}(X) =$
- Variance: $\text{Var}(X) =$

Hypergeometric: Hypergeometric(N, M, K): When drawing a sample of size K from a group of N items, M of which are special, 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: $\mathbb{E}(X) =$ (find by indicator method)



- Variance: $\mathrm{Var}(X) =$ (find by indicator method)

1.1.6 Indicator Method

Proposition 18. 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 19. *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 20. (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*

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

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

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

1.1.9 Asymptotic Distributions

Proposition 21.

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

Theorem 22. 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

Fall 2011 Problem 1 (same as HW1 problem 5; similar to HW3 problem 2(5); likely to be question 1 on the midterm.) True or false: if A and B are events such that $0 < \Pr(A) < 1$ and $\Pr(B \mid A) = \Pr(B \mid A^c)$, then A and B are independent.

Solution. 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 $\boxed{\text{true.}}$

Similar problem: 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

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

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

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

Fall 2014 Problem 1 (likely to be question 2 on the midterm). 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. ($\implies$) 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.

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.

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

(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

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

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

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

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

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

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

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

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

(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 $\{A_k|k = 1,2,\dots\} \cup \{B_k|k = 1,2,\dots\} \cup \{C_k|k = 1,2,\dots\}$. Since all of these events are mutually exclusive, we have

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

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

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

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

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

Similar: 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) $\mathrm{Cov}(\mathbb{E}(X \mid Y), Y) = \mathrm{Cov}(X, Y)$

(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)

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

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

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

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

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

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

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

Remaining problems are likely to be indicator method.

1.2.2 Problems we did in class that professor mentioned

Matching: n objects belong in n places. If placed randomly, what is the probability of at least one match?

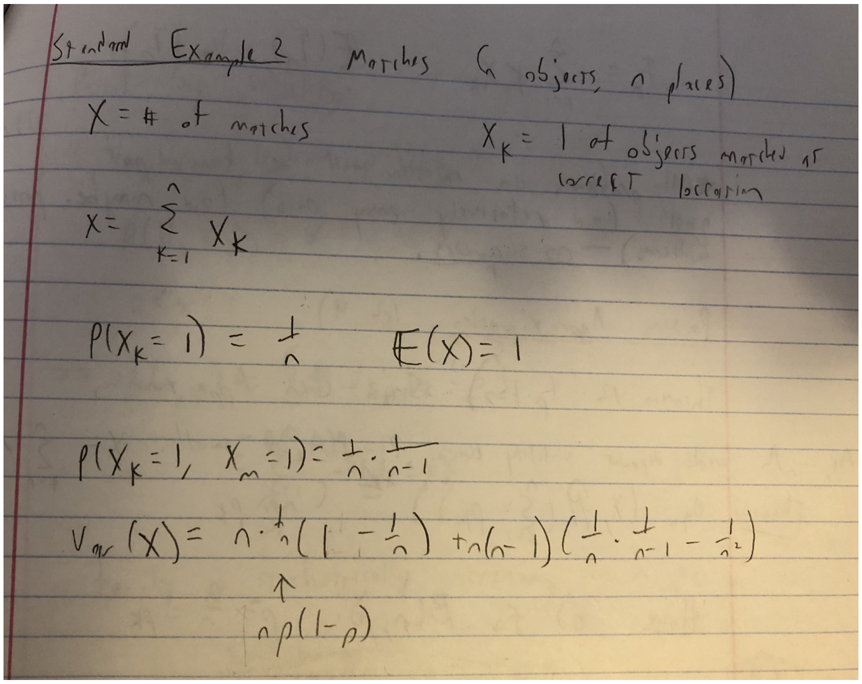
Let A_k = match for object k. Want: $P\left(\bigcup_{k=1}^{n} A\right)$

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

\# of groups of k objects that could be matched correctly

probability that k objects placed correctly (other $n-k$ placed randomly)

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



Standard Example 2 Matches (n objects, n places)

$X = \#$ of matches

$X_k = 1$ of objects matched at correct location

$$X = \sum_{k=1}^{n} X_k$$

$P(X_k = 1) = \frac{1}{n}$ $\mathbb{E}(X) = 1$

$P(X_k = 1, X_m = 1) = \frac{1}{n} \cdot \frac{1}{n-1}$

$$\mathrm{Var}(X) = \underbrace{n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right)}_{np(1-p)} + n(n-1)\left(\frac{1}{n} \cdot \frac{1}{n-1} - \frac{1}{n^2}\right)$$

Variance Problem 09/21 If $\mathbb{E}(X \mid Y) = Y, \mathbb{E}(Y \mid X) = X, \mathbb{E}(X^2) < \infty, \mathbb{E}(Y^2) < \infty$, show $\mathbb{E}(X - Y)^2 = 0$ (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,

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

Therefore

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

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 3 (HW3 Problem 8—almost full solution)

Problem 8. 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 one. Assume that N is independent of $U_1, U_2, \ldots$ and define

$$Y = \begin{cases} 0, & \text{if } N = 0, \\ \max_{1 \le i \le N} U_i, & \text{if } N > 0. \end{cases}$$

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 (Theorem 11),

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

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

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

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

$$\mathbb{E}(Y) = \mathbb{E}\left(\frac{N}{N+1}\right) = \sum_{n=1}^{\infty} \frac{n}{n+1} \Pr(N = n) = \sum_{n=1}^{\infty} \frac{n}{n+1} \frac{\exp(-1)1^n}{n!} = \boxed{\frac{1}{e} \sum_{n=1}^{\infty} \frac{n}{(n+1)!}}$$

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 $k \ge 1$ such that $k + 3 \le m$, 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

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

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

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

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

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

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

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

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

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

(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 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

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

begin by noting that we can also express N_m as

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

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 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

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

$$= 0\big) \approx \frac{\exp(-\lambda_{xm+m\log m}) \cdot \lambda^0_{xm+m\log m}}{0!} = \exp\big(-\lambda_{xm+m\log m}\big)$$

And we have

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

$$= \exp(-x)$$

which yields

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

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 $\boxed{f \text{ is a probability density function if } C = \frac{3}{4}.}$

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 9, 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 7 identical gifts to 10 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 $n = 7$ and $r = 10$. Therefore the number of possible allocations is

$$\binom{7+10-1}{10-1} = \boxed{11,440}$$

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

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

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

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.

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

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

So we have

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

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

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

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

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

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) Imagine we lay out the 20 indistinguishable gifts and take 6 sticks to place in the 19 spaces between them. The first child gets all the gifts to the left of the leftmost stick, the second child gets the gifts between the leftmost and second leftmost stick, and so on, until the last child gets all the gifts to the right of the rightmost stick. Because each child must receive at least one gift, we must place each stick in a different space. There are $\binom{19}{6} = \boxed{27,132}$ ways to do this. (See also Proposition 8.)

(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

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

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

$$x_1 - 1 + x_2 - 1 + \ldots + x_7 - 1 = 20 - 1 \cdot 7, x_i - 1 \geq 1 \; \forall i$$

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

$$y_1 + y_2 + \ldots + y_7 = 13, y_i \geq 1 \; \forall i$$

By Proposition 8 (and because of the same logic as used in part (a)), the number of distinct solutions to this equation, and therefore the number of possible allocations under these conditions, is $\binom{12}{6} = \boxed{924}$.

(c) By Proposition 9, 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

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

1.3 To Know for Math 505A Midterm 2

1.3.1 Definitions

Definition 1.15. 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 1.16. The function f is called the **(probability) density function** of the continuous random variable X.

Proposition 23. 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 1.17. 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 1.18. The random variables X and Y are **jointly continuous** with **joint (probability) density function** $f : \mathbb{R}^2 \to [0, \infty)$ if

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

Definition 1.19. 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 1.19: 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 24. 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 25. For continuous random variables, the previous condition is equivalent to requiring $f(x, y) = f_X(x)f_Y(y)$.
-

  Theorem 26. If two variables are bivariate normal, they are independent if and only if their covariance

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

  is equal to 0.
- Characteristic functions:

  Theorem 27. X and Y are independent if and only if $\phi_{X,Y}(s, t) = \phi_X(s)\phi_Y(t)$.

Theorem 28. (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 $h(Y)$.

1.3.2 Probability-Generating Functions

Definition 1.20.

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

Theorem 29. 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 1.21.

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

Theorem 30. 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 1.22.

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

Proposition 31. 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 32. 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 33. 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: 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:
- Characteristic function:

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

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

- Probability density function:

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

- 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: $\Gamma(\alpha, \beta)$ (**note: this parameterization is a little unusual; more commonly β is expressed as the reciprocal of how it appears here.**)

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

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:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = \beta = \lambda^{-1}$
- Variance: $\text{Var}(X) = \beta^2 = \lambda^{-2}$

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: $\text{Var}(X) =$

1.3.6 Multivariate Gaussian (Normal) Distributions

Definition 1.23. 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 35. From http://pluto.huji.ac.il/~pchiga/teaching/MathStat/SIAnotes2013.pdf (Proposition 3c1): Let X be a Gaussian random variable in $\mathbb{R}^2$ as in Definition 2b6. 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

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

Recall Theorem 26: if two variables are bivariate normal, they are independent if and only if their covariance

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

equals 0.

Theorem 36. For a bivariate normal distribution

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

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.

1.4 Worked problems

1.4.1 Example Problems That Will Likely Appear on Midterm

(1) **Note: we worked through an example problem like this on Friday. Should probably fix solution, and use geometric**

Question:

Let X, Y, Z be independent uniform on (0, 1). Compute the cdfs of XY, X/Y, and XY/Z.

Solution (may not be the way Lototsky suggested, consider revising).

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:

$$F_{XY}(z) = \int_0^\infty f_X(x) \int_{-\infty}^{z/x} f_Y(y)dydx - \int_{-\infty}^0 f_X(x) \int_\infty^{z/x} f_Y(y)dydx$$

$$= \int_0^1 \left[(z/x)\mathbf{1}_{\{0<z/x\leq 1\}} + \mathbf{1}_{\{z/x>1\}}\right]dx = \int_0^1 \left[(z/x)\mathbf{1}_{\{z\leq x\}} + \mathbf{1}_{\{z>x\}}\right]dx = \int_0^z dx + \int_z^1 (z/x)dx$$

$$= z + z\log(x)\Big|_z^1 = z + z\log(1) - z\log(z) = z(1-\log(z))$$

$$\implies \boxed{F_{XY}(z) = \begin{cases} 0 & z \leq 0 \\ z(1-\log(z)) & 0 < z \leq 1 \\ 1 & z > 1 \end{cases}}$$

- X/Y:

$$F_{X/Y}(z) = \int_0^\infty f_Y(y) \int_{-\infty}^{zy} f_X(x)dxdy - \int_{-\infty}^0 f_Y(y) \int_\infty^{zy} f_X(x)dxdy$$

$$= \int_0^1 \left[zy\mathbf{1}_{\{0<zy\leq 1\}} + \mathbf{1}_{\{zy>1\}}\right]dy = \int_0^1 \left[zy\mathbf{1}_{\{y>0\cap y\leq 1/z\}} + \mathbf{1}_{\{y>1/z\}}\right]dy = \int_0^{1/z} zy \cdot dy + \int_{1/z}^1 dy$$

$$= \frac{zy^2}{2}\Big|_0^{1/z} + (1-1/z) = \frac{z}{2z^2} + 1 - \frac{2}{2z} = 1 - \frac{1}{2z}$$

$$\implies F_{XY}(z) = \begin{cases} 0 & z \leq 0 \\ 1 - \frac{1}{2z} & 0 < 1/z \leq 1 \end{cases}$$

$$= \boxed{\begin{cases} 0 & z \leq 0 \\ z/2 & 0 < z \leq 1 \\ 1 - \frac{1}{2z} & z > 1 \end{cases}}$$

- XY/Z: Consider this the cdf of the quotient of $W = XY$ and Z.

$$F_U(u) = \int_0^\infty f_Z(z) \int_{-\infty}^{uz} f_W(w) dw dz - \int_{-\infty}^{0} f_Z(z) \int_{\infty}^{uz} f_W(w) dw dz$$

$$= \int_0^1 \int_0^{uz} -\log(w) \mathbf{1}_{\{0 < uz \leq 1\}} dw dz = \int_0^1 -\big[w \log(w) - w\big]_0^{uz} \mathbf{1}_{\{0 < z \leq 1/u\}} dz$$

$$= \int_0^{1/u} uz\big[1 - \log(uz)\big] dz = \frac{u}{4} z^2 \big(3 - 2\log(uz)\big)\Big|_0^{1/u} = \frac{u}{4u^2}\big(3 - 2\log(1)\big) - 0 = \frac{3}{4u}$$

$$\implies \boxed{F_{XY/Z}(u) = \begin{cases} 0 & u \leq 0 \\ \frac{3}{4u} & 0 < u \leq 3/4 \\ 1 & u > 3/4 \end{cases}}$$

(2) **Note: we worked through an example problem like this on Friday.**

Question from Friday: Let X, Y be distributed exponentially with mean 1. What is the probability distribution of $X/(X+Y)$?

Solution. Find the cdf:

$$\Pr\left(\frac{X}{X+Y} < t\right) = \Pr(X < tX + tY) = \Pr\left(Y > \frac{X(1-t)}{t}\right)$$

To find $\Pr(Y > aX)$, graph the line aX

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

- $\mathbb{E}(S) = M_S'(0) = G_N'\big(M_X(0)\big) M_X'(0) = G_N'(1)\mathbb{E}(X_1) = \mathbb{E}(N)\mathbb{E}(X_1) = \mu_N \mu_X$
- $\text{Var}(S) = \mathbb{E}(S^2) - \mathbb{E}(S)^2 = M_S''(0) - (M_S'(0))^2$

$$= G_N''\big(M_X(0)\big)(M_X'(0))^2 + G_N'\big(M_X(0)\big) M_X''(0) - \mu_N^2 \mu_X^2 = G_N''\big(1\big)\mathbb{E}(X_1)^2 + G_N'\big(1\big)\text{Var}(X_1) - \mu_N^2 \mu_X^2$$

$$= \mathbb{E}[N(N-1)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2 \mu_X^2 = \mathbb{E}[N^2 - N]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2 \mu_X^2$$

$$= [\mathbb{E}(N^2) - \mathbb{E}(N)^2 + \mathbb{E}(N)^2 - \mathbb{E}(N)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2 \mu_X^2 =$$

$$= [\mathrm{Var}(N) + \mathbb{E}(N)^2 - \mathbb{E}(N)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\mathrm{Var}(X_1) - \mu_N^2\mu_X^2 = (\sigma_N^2 + \mu_N^2 - \mu_N)\mu_X^2 + \mu_N\sigma_X^2 - \mu_N^2\mu_X^2$$

$$= \boxed{(\sigma_N^2 - \mu_N)\mu_X^2 + \mu_N\sigma_X^2}$$

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

$$f(x) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{x-1}{365}\right) \approx \exp(-1/365)\exp(-2/365) \cdots \exp(-(x-1)/365)$$

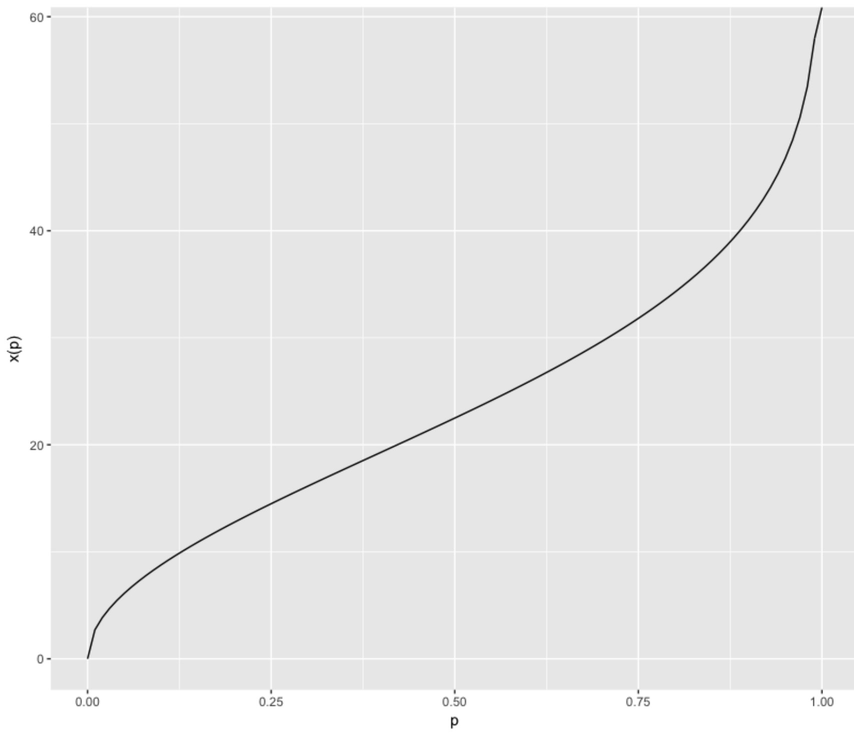
$$= e^{-(x^2 - x)/(2 \cdot 365)}$$

We want the probability of a match to be at least p; that is, $f(x) \leq 1 - p$. 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 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.