Kyle's Probability Theory

Cody Vig

December 2021 — January 2022

Problem Set 1: Combinatorics

1. A finite set Ω has n elements. Show that if we count the empty set and Ω as subsets, there are 2^n subsets of Ω.

   I will provide two proofs; one which uses the binomial theorem, and one which uses induction. The proof using the binomial theorem is as follows.

   Proof. Note that the number of elements of each subset must range between 0 (i.e., the empty set) and n (i.e., Ω). Hence, we can count the total number of subsets by counting the total number of subsets of size k for $0 \le k \le n$ and summing over k. Since there are $\binom{n}{k}$ many subsets of size k, we find

   $$\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} 1^k \cdot 1^{n-k} = (1+1)^n = 2^n$$

   where in the second to last equality, I have used the binomial theorem. This completes the proof. $\square$

   The second proof uses induction on the cardinality of Ω.

   Proof. Suppose $|\Omega| = 0$, i.e., $\Omega = \varnothing$. Then the set of subsets of Ω is $\{\varnothing\}$, which has $1 = 2^0$ elements. The completes the base-case. Next, assume that $\Omega = \{\omega_1, \ldots, \omega_n\}$ has $n > 0$ elements and that the set of all subsets of Ω has 2^n elements. Consider the set

   $$\Omega' = \Omega \cup \{\omega_{n+1}\} = \{\omega_1, \ldots, \omega_{n+1}\}.$$

   Either a given subset of Ω' contains ω_{n+1} or it does not. If a given subset does not contain ω_{n+1}, then it is a subset of Ω, of which there are 2^n. If a given set subset does contain ω_{n+1}, then we can write such a subset as $\{\omega_{n+1}\} \cup A$ where $A \subseteq \Omega$; hence choosing such a subset which contains ω_{n+1} is equivalent to choosing a subset of Ω, of which there are 2_n many choices. Since every subset of Ω' either contains ω_{n+1} or it does not, we have

   $$2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$$

   many subsets of Ω'. This completes the induction. $\square$

2. If a set has $2n$ elements, show that it has more subsets with n elements than with any other number of elements.

   Solution: Since there are precisely $\binom{2n}{k}$ many subsets of size k, we equivalently want to show:

   $$\binom{2n}{n} \ge \binom{2n}{k} \iff \frac{(2n)!}{(n!)^2} \ge \frac{(2n)!}{k!(2n-k)!} \iff \frac{k!(2n-k)!}{(n!)^2} \ge 1.$$

Assume without loss of generality that $k \leq n$ (if $k > n$, then replace $k \mapsto 2n - k \leq n$) so that $2n - k \geq n \geq k$. Write

$$\frac{k!(2n-k)!}{(n!)^2} = \frac{(2n-k)!/n!}{n!/k!} = \frac{2n-k}{n} \cdot \frac{2n-k-1}{n-1} \cdots \frac{n+1}{k+1}.$$

In particular, since there are $(2n-k) - n = n - k$ factors in the numerator and $n - k$ factors in the denominator, such a decomposition of the numerator and denominator is possible. Since $2n - k \geq n$, $2n - k - \ell \geq n - \ell$ for all $0 \leq \ell \leq 2n - 2k - 1$, and so each constituent factor is greater than or equal to 1, with equality holding only when $k = n$. Hence, since the product of numbers bigger than or equal to 1 is bigger than or equal to 1, we find $k!(2n-k)!/(n!)^2 \geq 1$, which completes the proof.

3. Prove the binomial theorem. (Prefer using a combinatorics approach)

First, we need a lemma.

Lemma 1. *If $m > 1$ and $0 \leq k < m$, then $\binom{m}{k} = \binom{m-1}{k-1} + \binom{m-1}{k}$.*

Proof. I will prove the lemma by counting the number of k-element subsets of a set A of size m in two different ways. On the one hand, there are clearly $\binom{m}{k}$ many such subsets. On the other hand, consider a fixed element a in A; either a given k-element subset contains a or it does not. If it does contain a, then it may also contain any $k-1$ elements from the remaining $m-1$ elements of A, and so there are $\binom{m-1}{k-1}$ many such subsets which do contain a. If the subset does not contain a, then it must contain exactly k elements of the remaining $m-1$ elements of A, and so there are $\binom{m-1}{k}$ many subsets which do not contain A. Hence, there are $\binom{m-1}{k-1} + \binom{m-1}{k}$ subsets of size k. This completes the proof. □

The binomial theorem states that if x and y are real numbers and n is a positive integer, then

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

where $\binom{n}{k} := n!/(k!(n-k)!)$ is the binomial coefficient. I will provide two proofs; one which uses induction on n, and one which uses a combinatorical approach. The proof using induction is as follows.

Proof. Observe $(x+y)^1 = x + y = \binom{1}{0}x^0y^{1-0} + \binom{1}{1}x^1y^{1-1}$, so the statement is true for $n = 1$. Next assume that the binomial theorem holds for $n = m - 1$ for $m > 1$. I will show that this implies the binomial theorem holds for $n = m$. Note:

$$\begin{aligned}(x+y)^m = (x+y)(x+y)^{m-1} &= (x+y)\sum_{k=0}^{m-1}\binom{m-1}{k}x^k y^{m-(k+1)} \\ &= \sum_{k=0}^{m-1}\binom{m-1}{k}x^{k+1}y^{m-(k+1)} + \sum_{k=0}^{m-1}\binom{m-1}{k}x^k y^{m-k}\end{aligned}$$

where the penultimate equality uses the inductive hypothesis. Next, we re-index the first sum, raplacing $k \mapsto k+1$. Then:

$$(x+y)^m = \sum_{k=0}^{m-1}\binom{m-1}{k}x^{k+1}y^{m-(k+1)} + \sum_{k=0}^{m-1}\binom{m-1}{k}x^k y^{m-k}$$

$$= \sum_{k=1}^{m} \binom{m-1}{k-1} x^k y^{m-k} + \sum_{k=0}^{m-1} \binom{m-1}{k} x^k y^{m-k}$$
$$= \binom{m}{m} x^m + \binom{m-1}{0} y^m + \sum_{k=1}^{m-1} \left\{ \binom{m-1}{k-1} + \binom{m-1}{k} \right\} x^k y^{m-k}.$$

Note that $\binom{m}{m} = 1 = \binom{m-1}{0}$. Further, from the lemma, we get

$$(x+y)^m = x^m + y^m + \sum_{k=1}^{m-1} \binom{m}{k} x^k y^{m-k} = \sum_{k=0}^{m} \binom{m}{k} x^k y^{m-k}.$$

This completes the proof □

The next proof is combinatorical.

Proof. Write

$$(x+y)^n = \underbrace{(x+y)(x+y)\cdots(x+y)}_{n}.$$

Note that the product will be a polynomial whose terms are of the form $c_k x^{a_k} y^{b_k}$ where a_k, b_k, and c_k are nonnegative integers for each $0 \leq k \leq n$. Further, since there are n factors of the form $(x+y)$ and each term can contain precisely one of the x's or y's from each factor, we must have $b_k + c_k = n$. In particular, the k'th term will be $c_k x^k y^{n-k}$. To determine c_k, we count the number of ways to select k x's from the product. Since there are n factors and each factor can either give an x or give a y (i.e., *not* an x), there are $c_k = \binom{n}{k}$ many ways to select k x's. Since the product can be expanded into the sum of all the terms $c_k x^k y^{n-k}$, we find

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},$$

which is what we wanted to prove. □

Problem Set 2: Axiomatic Probability Theory

For the sake of clarity, the probability axioms are included below. We define Ω to be a sample space and $P : \Omega \to \mathbb{R}$ to be a function satisfying the following three axioms.

- (Pr_1): For any event $A \in \Omega$, $\mathbb{P}(A) \geq 0$;
- (Pr_2): $\mathbb{P}(\Omega) = 1$;
- (Pr_3): If $\{A_i\}_{i=1}^{\infty} \subset \Omega$ are mutually exclusive (i.e., $A_i \cap A_j = \varnothing$ for all $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

1. Basic properties of the the probability function.

   (a) Show that $\mathbb{P}(A^{\mathsf{c}}) = 1 - \mathbb{P}(A)$.

   Solution: Note that $A \cup A^{\mathsf{c}} = \Omega$ and $A \cap A^{\mathsf{c}} = \varnothing$. By ($\mathrm{Pr}_2$) and ($\mathrm{Pr}_3$), we have $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^{\mathsf{c}}) = \mathbb{P}(A) + \mathbb{P}(A^{\mathsf{c}})$, and so $\mathbb{P}(A^{\mathsf{c}}) = 1 - \mathbb{P}(A)$.

   (b) Show that $\mathbb{P}(\varnothing) = 0$.

   Solution: Since $\varnothing = \Omega^{\mathsf{c}}$, ($\mathrm{Pr}_2$) and Problem 1a imply that $\mathbb{P}(\varnothing) = \mathbb{P}(\Omega^{\mathsf{c}}) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$.

   (c) Suppose that A is a subset of B, which are events of some sample space. Show that $\mathbb{P}(A) \leq \mathbb{P}(B)$.

   Solution: Let $C = B \setminus A$. Then by construction, $A \cap C = \varnothing$ and $A \cup C = B$. From (Pr_3), we get $\mathbb{P}(B) = \mathbb{P}(A \cup C) = \mathbb{P}(A) + \mathbb{P}(C) \geq \mathbb{P}(A)$ since $\mathbb{P}(C) \geq 0$ by (Pr_1).

   (d) Show that if A is an event then $\mathbb{P}(A) \leq 1$.

   Solution: For any $A \in \Omega$, $A \subseteq \Omega$. From (Pr_2) and Problem 1c, we find $\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$. This shows that the statement "For any $A \in \Omega$, $0 \leq \mathbb{P}(A) \leq 1$" is a stronger axiom than necessary, which is why it has been replaced with (Pr_1).

2. Some fundamental theorems of probability.

   (a) Prove Bayes' Theorem.

   Solution: Let A and B be events in a sample space Ω with $\mathbb{P}(B) \neq 0$. Then Bayes' Theorem states that $\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A)\mathbb{P}(A)/\mathbb{P}(B)$. Recall that $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ by definition. Then

$$\mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) = \mathbb{P}(B \mid A)\mathbb{P}(A).$$

   Bayes' theorem follows.

   (b) Prove the Law of Total Probability.

   Solution: Let $\{B_n\}$ be a (possibly countably infinite) subset of Ω such that

$$\bigcup_{n=1}^{\infty} B_n = \Omega \quad \text{and} \quad B_i \cap B_j = \varnothing \quad \forall i \neq j.$$

(Such a set is called a *partition* of Ω and is usually denoted by the disjoint union $\sqcup_n B_n = \Omega$.) Then the law of total probability is that for any $A \in \Omega$,

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A \cap B_n).$$

Proof. Note that $\cup_{n=1}^{\infty}(A \cap B_n) = A \cap (\cup_{n=1}^{\infty} B_n) = A \cap \Omega = A$. Since $B_i \cap B_j = \varnothing$ for all $i \neq j$, we have $(A \cap B_i) \cap (A \cap B_j) = \varnothing$ for all $i \neq j$, and so

$$\mathbb{P}(A) = P\left(\bigcup_{n=1}^{\infty}(A \cap B_n)\right) = \sum_{n=1}^{\infty} \mathbb{P}(A \cap B_n),$$

which completes the proof. □

3. Prove that if A and B are mutually exclusive, then $\mathbb{P}(A \mid A \text{ or } B) = \mathbb{P}(A)/(\mathbb{P}(A) + \mathbb{P}(B))$.

   Solution: If C and D are events in a sample space, then we have $\mathbb{P}(C \mid D) := \mathbb{P}(C \cap D)/\mathbb{P}(D)$ by definition. Hence:

   $$\mathbb{P}(A \mid A \cup B) = \frac{\mathbb{P}(A \cap (A \cup B))}{\mathbb{P}(A \cup B)}.$$

   Since A and B are mutually exclusive, $\mathbb{P}(A \cap B) = 0$. From the result of Problem 8, we have $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$. Additionally, $A \cap (A \cup B) = A$ since $A \subseteq (A \cup B)$. Thus,

   $$\mathbb{P}(A \mid A \cup B) = \frac{\mathbb{P}(A \cap (A \cup B))}{\mathbb{P}(A \cup B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A) + \mathbb{P}(B)},$$

   which completes the proof.

4. Prove the Inclusion-Exclusion Identity for 2 events. ($\mathbb{P}(A \text{ or } B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \text{ and } B)$)

   Solution: Recall that if $A \in \Omega$, then $\mathbb{P}(A) = |A|/|\Omega|$, where $|\cdot|$ denotes cardinality. Hence, the claim is equivalent to

   $$\begin{aligned} & \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ \iff \quad & |A \cup B| = |A| + |B| - |\mathbb{P}(A \cap B)|. \end{aligned}$$

   This is the statement I will prove.[1] Write

   $$\begin{aligned} A &= (A \setminus B) \cup (A \cap B) \\ B &= (B \setminus A) \cup (A \cap B) \end{aligned}$$

   Then, since $A \setminus B$ and $A \cap B$ are disjoint by construction, we have $|A| = |A \setminus B| + |A \cap B|$. Similarly, $|B| = |B \setminus A| + |A \cap B|$. Finally, we get

   $$\begin{aligned} |A| + |B| - |A \cap B| &= |A \setminus B| + |B \setminus A| + |A \cap B| \\ &= |(A \setminus B) \cup (B \setminus A)| + |A \cap B| \\ &= |(A \cup B) \setminus (A \cap B)| + |A \cap B| \\ &= |A \cup B| - |A \cap B| + |A \cap B| \\ &= |A \cup B|, \end{aligned}$$

   which completes the proof.

[1] This might not be sufficiently general. Nothing in the probability axioms specify that P must be a ratio of set sizes. A more general proof is necessary for the more pathological cases.

5. If A and B are independent, show that $\mathbb{P}(A^{\mathsf{c}} \text{ and } B^{\mathsf{c}}) = \mathbb{P}(A^{\mathsf{c}})\mathbb{P}(B^{\mathsf{c}})$.

   Solution: By DeMorgan's Law, $A^{\mathsf{c}} \cap B^{\mathsf{c}} = (A \cup B)^{\mathsf{c}}$. Hence

$$\mathbb{P}(A^{\mathsf{c}} \cap B^{\mathsf{c}}) = \mathbb{P}((A \cup B)^{\mathsf{c}}) = 1 - \mathbb{P}(A \cup B) = 1 - [\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)],$$

   where the last equality follows from the inclusion-exclusion identity (Problem 4). Since A and B are independent, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. So:

$$\mathbb{P}(A^{\mathsf{c}} \cap B^{\mathsf{c}}) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = \mathbb{P}(A^{\mathsf{c}})\mathbb{P}(B^{\mathsf{c}}).$$

   Thus if A and B are independent, so too are A^{c} and B^{c}.

Problem Set 3: Expectation and Variance

1. For real numbers a and b and $X : \Omega \to \mathbb{R}$ a random variable, prove $\mathrm{E}(aX + b) = aE(X) + b$.

   Solution: In this case, we find the expected value to be

$$\begin{aligned}\mathrm{E}(aX+b) &= \sum_x (ax+b)\mathbb{P}(aX+b = ax+b)\\ &= a\sum_x x\mathbb{P}(aX+b=ax+b) + b\sum_x \mathbb{P}(aX+b=ax+b)\\ &= a\sum_x x\mathbb{P}(aX+b=ax+b)+b,\end{aligned}$$

   since $\sum_x \mathbb{P}(aX+b=ax+b)=1$. Additionally, since $x \mapsto ax+b$ is a bijective transformation, $\mathbb{P}(aX+b=ax+b) = \mathbb{P}(X=x)$, and so

$$\mathrm{E}(aX+b) = a\sum_x x\mathbb{P}(X=x)+b = a\mathrm{E}(X)+b.$$

   This completes the proof.

2. Prove that if X and Y are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)$ and $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

   Solution: I will prove this for the case of discrete distributions, but the proof is easily adapted for the continuous case. Throughout the problem, we define the probability mass functions $p_X(x) := \mathbb{P}_X(X=x)$ and $p_Y(y) := \mathbb{P}_Y(Y=y)$.

   Claim 1. *If X and Y are independent, then $E(XY) = E(X)E(Y)$.*

   Proof. Since X and Y are independent, we have $\mathbb{P}_{XY}(X=x, Y=y) = p_X(x)p_Y(y)$. Then,

$$\begin{aligned}\mathrm{E}(XY) = \sum_x\sum_y xy\mathbb{P}_{XY}(X=x,Y=y) &= \sum_x\sum_y xyp_X(x)p_Y(y)\\ &= \sum_x xp_X(x)\sum_y yp_Y(y) = \mathrm{E}(X)\mathrm{E}(Y),\end{aligned}$$

   as expected. □

   The following lemma will be used to prove the linearity of variance.

   Lemma 2. *Let X and Y be (not necessarily independent) random variables. Then $E(X+Y) = E(X)+E(Y)$.*

   Proof. We start with the right-hand side of the equation and derive the right using the notion of marginal distributions. So,

$$\begin{aligned}\mathrm{E}(X)+\mathrm{E}(Y) &= \sum_x x\mathbb{P}_X(X=x) + \sum_y y\mathbb{P}_Y(Y=y)\\ &= \sum_x\sum_y x\mathbb{P}_{XY}(X=x,Y=y) + \sum_x\sum_y y\mathbb{P}_{XY}(X=x,Y=y)\\ &= \sum_x\sum_y (x+y)\mathbb{P}_{XY}(X=x,Y=y)\\ &= \mathrm{E}(X+Y).\end{aligned}$$

   This, together with Problem 1, proves the linearity of the expected value. □

Claim 2. *If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$.*

Proof. Using the results of part (a) and the lemma above, we find

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}([X+Y]^2) - \mathrm{E}^2(X+Y) \\
&= [\mathrm{E}(X^2) + 2\mathrm{E}(X)\mathrm{E}(Y) + \mathrm{E}(Y^2)] - [\mathrm{E}^2(X) + 2\mathrm{E}(x)\mathrm{E}(Y) + \mathrm{E}^2(Y)] \\
&= \mathrm{E}(X^2) - \mathrm{E}^2(X) + \mathrm{E}(Y^2) - \mathrm{E}^2(Y) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y),
\end{aligned}$$

as expected. □

3. Prove the Law of the Unconscious Statistician.

Solution: For any real-valued function g, the law of the unconscious statistician states that

$$\mathrm{E}(g(X)) = \sum_n g(x_n)\mathbb{P}(X = x_n).$$

By definition, we have $\mathrm{E}(g(X)) = \sum_n y_n \mathbb{P}(g(X) = y_n)$ where $y_n = g(x_n)$. We convert to a sum over the $g(x_n)$ by grouping all x_{n_k} such that $g(x_{n_k}) = y_n$. Let I_n be the index set for $g^{-1}(y_n)$; that is, $I_n := \{m \mid x_m \in g^{-1}(y_n)\}$. Then, we write

$$\begin{aligned}
\mathrm{E}(g(X)) = \sum_n y_n \mathbb{P}(g(X) = y_n) &= \sum_m g(x_m) \sum_{p \in I_n} \mathbb{P}(X = x_m) \\
&= \sum_m \sum_{p \in I_n} g(x_m)\mathbb{P}(X = x_m) = \sum_q g(x_q)\mathbb{P}(X = x_q).
\end{aligned}$$

In the second step, I used the fact that the probability that the image of X is y_n is just the probability that X is one of the elements of $g^{-1}(y_n)$. This completes the proof.

4. Show that $\mathrm{Var}(X) = 0$ if and only if there is a constant c such that $\mathbb{P}(X = c) = 1$.

Solution: For simplicity, define $p(x) := \mathbb{P}(X = x)$ to be the probability mass function for X.

$(\implies)$ Suppose $\mathrm{Var}(X) = 0$. Then

$$0 = \mathrm{Var}(X) = \mathrm{E}([X - \mathrm{E}(X)]^2) = \sum_n (x_n - \mathrm{E}(X))^2 p(x_n).$$

Since both $(x - \mathrm{E}(X))^2 \geq 0$ and $p(x) \geq 0$ for all x, we must have

$$(x_n - \mathrm{E}(X))^2 p(x_n) = 0$$

for each n. This either implies $p(x_n) = 0$ or that $x_n = \mathrm{E}(X)$. Since $\sum_n p(x_n) = 1$, there must be a countable subset $\{x_{n_k}\}$ for which $p(x_{n_k}) \neq 0$ and $x_{n_k} = \mathrm{E}(X)$. However, if each outcome is equal to the expected outcome, all of the outcomes are the same, and so there exists a c such that $x_{n_1} = x_{n_2} = \cdots = c$, which implies $p(c) = 1$.

$(\impliedby)$ Suppose there exists a c such that $\mathbb{P}(c) = 1$. Then since $\sum_x \mathbb{P}(x) = 1$, we have $\mathbb{P}(x) = 0$ for all $x \neq c$. Then

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \sum_x x^2\mathbb{P}(x) + \Big(\sum_x x\mathbb{P}(x)\Big)^2 = c^2\mathbb{P}(c) - (c\mathbb{P}(c))^2 = c^2 - c^2 = 0$$

since $\mathbb{P}(c) = 1$.

Problem Set 4: Discrete Distributions

1. Prove the Expected Value and Variance of your favorite 2 distributions in the following:

(a) Hypergeometric Distribution

Solution: Coming soon!

(b) Negative Binomial Distribution

Solution: Coming soon!

(c) Poisson Distribution

Solution: Let $X \sim \text{Poisson}(\lambda)$. Then $\mathbb{P}(X = n) = e^{-\lambda}\lambda^n/n!$. We have

$$\mathrm{E}(X) = \sum_{n=0}^{\infty} n e^{-\lambda}\frac{\lambda^n}{n!} = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

Additionally, using the law of the unconscious statistician, we get

$$\begin{aligned}\mathrm{E}(X^2) = \sum_{n=0}^{\infty} n^2 e^{-\lambda}\frac{\lambda^n}{n!} = \lambda e^{-\lambda}\sum_{n=1}^{\infty}\frac{n}{(n-1)!}\lambda^{n-1} &= \lambda e^{-\lambda}\sum_{n=0}^{\infty}\frac{n+1}{n!}\lambda^n \\ &= \lambda e^{-\lambda}(\lambda+1)\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = \lambda e^{-\lambda}(\lambda+1)e^{\lambda} = \lambda^2 + \lambda.\end{aligned}$$

Hence the variance is $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

(d) Uniform Distribution

Solution: Without loss of generality, we consider a uniform distribution $\mathrm{U}(1, N)$ on the set $\mathbb{I}_N := \{1, 2, \ldots, N\}$ for some positive integer N, since any other countable set can be bijectively mapped onto $\mathbb{I}_N$. Then if $X \sim \mathrm{U}(1, N)$, we have $\mathbb{P}(X = k) = 1/N$ for each k in $\mathbb{I}_N$. Then

$$\mathrm{E}(X) = \sum_{k=1}^{N} k\mathbb{P}(X = k) = \frac{1}{N}\sum_{k=1}^{N} k = \frac{N(N+1)}{2N} = \frac{N+1}{2}.$$

Also

$$\mathrm{E}(X^2) = \sum_{k=1}^{N} k^2\mathbb{P}(X = k) = \frac{1}{N}\sum_{k=1}^{N} k^2 = \frac{N(N+1)(2N+1)}{6N} = \frac{(N+1)(2N+1)}{6}.$$

Hence, the variance is

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} = \frac{(N+1)(N-1)}{12} = \frac{N^2-1}{12}.$$

We can easily generalize to $\mathrm{U}(a, b)$ by writing $\mathrm{U}(a, b) = \mathrm{U}(0, b) \setminus \mathrm{U}(0, a-1)$.

(e) Binomial Distribution

Solution: A binomial distribution B is just a sequence of n independent Bernoulli trials (i.e., indicator random variables I). These are known to have $\mathrm{E}(I) = p$ and $\mathrm{Var}(I) = pq$, where $q = 1 - p$. From Problem 1 of Problem Set 3, we have $\mathrm{E}(B) = \mathrm{E}(nI) = n\mathrm{E}(I) = np$, and from Problem 2 of Problem Set 3, we have $\mathrm{Var}(B) = \mathrm{Var}(nI) = n\mathrm{Var}(I) = npq$.

(f) Geometric Distribution
Solution: Let $X \sim \text{Geometric}(p)$. Then $\mathbb{P}(X = n) = (1-p)^{n-1}p = q^{n-1}p$. The expected value is therefore:

$$\mathrm{E}(X) = \sum_{n=1}^{\infty} nq^{n-1}p = \sum_{n=0}^{\infty}(n+1)q^n p = q\sum_{n=1}^{\infty} nq^{n-1}p + \sum_{n=1}^{\infty} q^{n-1}p = q\mathrm{E}(X) + 1,$$

hence $1 = (1-q)\mathrm{E}(X) = p\mathrm{E}(x)$, and so $\mathrm{E}(X) = 1/p$. Then

$$\begin{aligned}\mathrm{E}(X^2) &= \sum_{n=1}^{\infty} n^2 q^{n-1}p = \sum_{n=0}^{\infty}(n+1)^2 q^n p = \sum_{n=0}^{\infty} n^2 q^n p + 2\sum_{n=0}^{\infty} nq^n p + \sum_{n=0}^{\infty} q^n p \\ &= q\sum_{n=1}^{\infty} n^2 q^{n-1}p + 2q\sum_{n=1}^{\infty} nq^{n-1}p + \sum_{n=0}^{\infty} q^n p = q\mathrm{E}(X^2) + 2q\mathrm{E}(X) + 1,\end{aligned}$$

hence $(1-q)\mathrm{E}(X^2) = p\mathrm{E}(X^2) = 2q\mathrm{E}(X) + 1 = 2q/p + 1$, and so $\mathrm{E}(X^2) = (2-p)/p^2$. Finally, we get

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

as our variance.

2. Suppose that X has Poisson (λ) distribution, and that Y has geometric (p) distribution on $\{0, 1, 2, \ldots\}$ independently of X.

(a) Find a formula for $\mathbb{P}(Y \geq X)$ in terms of p and λ.
Solution: The key insight is observing that

$$\mathbb{P}(Y \geq X) = \sum_{n=0}^{\infty} \mathbb{P}(\{X = n\} \cap \{Y \geq n\}) = \sum_{n=0}^{\infty} \mathbb{P}(X = n)\mathbb{P}(Y \geq n),$$

since X and Y are independent. Since $X \sim \text{Poisson}(\lambda)$, we have $\mathbb{P}(X = n) = e^{-\lambda}\lambda^n/n!$; and since $Y \sim \text{Geometric}(p)$, we have $\mathbb{P}(Y = n) = q^n p$, and so $\mathbb{P}(Y \geq n) = \sum_{m=n}^{\infty} q^m p$. So:

$$\begin{aligned}\mathbb{P}(Y \geq X) &= \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\left\{\sum_{m=n}^{\infty} q^m p\right\} = \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\left\{1 - \sum_{m=0}^{n-1} q^m p\right\} \\ &= \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!}\left\{1 - \frac{\mathbb{P}(1-q^n)}{1-q}\right\} = \sum_{n=0}^{\infty} e^{-\lambda}\frac{(\lambda q)^n}{n!} = e^{\lambda(q-1)} = e^{-\lambda p}.\end{aligned}$$

Hence $\mathbb{P}(Y \geq X) = e^{-\lambda p}$.

(b) Evaluate numerically for $p = 1/2$ and $\lambda = 1$.
Solution: For $p = 1/2$ and $\lambda = 1$, we get $\mathbb{P}(Y \geq X) = e^{-1/2} \approx 60.7\%$.

3. Suppose we toss a coin once and let p be the probability of heads. Let X denote the number of heads and let Y denote the number of tails.

(a) Prove that X and Y are dependent.

Solution: Since there are only two outcomes for each coin toss and each subsequent toss is independent of the previous, the distribution is binomial with probability p of success (heads) and $q = 1 - p$ of failure (tails). If we toss the coin n times, then:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k} = \mathbb{P}(Y = n - k)$$

In particular, $X + Y = n$, and so $Y = n - X$. Each of these random variables is a function of the other, so if one is know, the other is known. Hence they are dependent. As an explicit counterexample, suppose we are flipping a coin three times. Clearly, say, $\mathbb{P}(\{X = 3\} \cap \{Y = 3\}) = 0$ since we would need at six coin tosses for there to be three heads *and* three tails. However,

$$\mathbb{P}(X = 3) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \frac{1}{8} = \binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \mathbb{P}(Y = 3),$$

and so $\mathbb{P}(X = 3)\mathbb{P}(Y = 3) = 1/8^2$ which is not zero, and so

$$\mathbb{P}(\{X = 3\} \cap \{Y = 3\}) \neq \mathbb{P}(X = 3)\mathbb{P}(Y = 3).$$

This proves that X and Y are not independent.

(b) Let $N \sim \text{Poisson}(\lambda)$ and suppose we toss a coin N times. Let X and Y be the number of heads and tails. Show that X and Y are independent.

Solution: It is sufficient to show $\mathbb{P}_{XY}(X = x, Y = y) = \mathbb{P}_X(X = x)\mathbb{P}_Y(Y = y)$. If $N \sim \text{Poisson}(\lambda)$, then $\mathbb{P}_N(N = n) = e^{-\lambda}\frac{\lambda^n}{n!}$. Constructing the marginal distribution, we find:

$$\mathbb{P}_X(X = k) = \sum_{n=k}^{\infty} \mathbb{P}_{XN}(X = k, N = n) = \sum_{n=k}^{\infty} \mathbb{P}_X(X = k \mid N = n)\mathbb{P}_N(N = n).$$

For a fixed n, $X \sim \text{Binomial}(n, p)$, so $\mathbb{P}_X(X = k \mid N = n) = \binom{n}{k}p^k q^{n-k}$. Then since N is Poisson distributed, we have

$$\begin{aligned}\mathbb{P}_X(X = k) = \sum_{n=k}^{\infty} \mathbb{P}_X(X = k \mid N = n)\mathbb{P}_N(N = n) &= \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} e^{-\lambda}\frac{\lambda^n}{n!} \\ &= \frac{1}{k!}\left(\frac{p}{q}\right)^k e^{-\lambda} \sum_{n=k}^{\infty} \frac{(\lambda q)^n}{(n-k)!} = \frac{(\lambda p)^k}{k!} e^{-\lambda} \underbrace{\sum_{n=0}^{\infty} \frac{(\lambda q)^n}{n!}}_{e^{\lambda q}} = \frac{(\lambda p)^k}{k!} e^{-\lambda p}.\end{aligned}$$

By symmetry then, we have:

$$\begin{aligned}\mathbb{P}_X(X = k) &= \frac{(\lambda p)^k}{k!} e^{-\lambda p} \\ \mathbb{P}_Y(Y = \ell) &= \frac{(\lambda q)^\ell}{\ell!} e^{-\lambda q}.\end{aligned}$$

In this case:

$$\mathbb{P}_X(X = k)\mathbb{P}_Y(Y = \ell) = \frac{(\lambda p)^k}{k!}\frac{(\lambda q)^\ell}{\ell!} e^{-\lambda}.$$

On the other hand, we may compute $\mathbb{P}_{XY}(X = k, Y = \ell)$ in the same way. We have

$$\begin{aligned}\mathbb{P}_{XY}(X = k, Y = \ell) &= \sum_{n=\min\{k,\ell\}}^{\infty} \mathbb{P}_{XYN}(X = k, Y = \ell, N = n) \\ &= \sum_{n=\min\{k,\ell\}}^{\infty} \mathbb{P}_{XY}(X = k, Y = \ell \mid N = n)\mathbb{P}_N(N = n).\end{aligned}$$

Since $n = k + \ell$ is the only n for which $X = k$ and $Y = \ell$, the sum collapses there and we get

$$\begin{aligned}\mathbb{P}_{XY}(X = k, Y = \ell) &= \mathbb{P}_{XY}(X = k, Y = \ell \mid N = k + \ell)\mathbb{P}_N(N = k + \ell) \\ &= \mathbb{P}_X(X = k \mid N = k + \ell)\mathbb{P}_N(N = k + \ell) \\ &= \binom{k+\ell}{k} p^k q^\ell e^{-\lambda} \frac{\lambda^{k+\ell}}{(k+\ell)!} \\ &= \frac{(\lambda p)^k}{k!}\frac{(\lambda q)^\ell}{\ell!} e^{-\lambda} \\ &= \mathbb{P}_X(X = k)\mathbb{P}_Y(Y = \ell),\end{aligned}$$

hence X and Y are indeed independent.

4. Prove the following properties about the CDFs:

(a) F is a nondecreasing function; that is, if $a < b$, then $F(a) \leq F(b)$

Solution: For a particular random variable X we define the cumulative distribution function (CDF) $F : \mathbb{R} \to \mathbb{R}$ by

$$F(x) = \mathbb{P}(X \leq x) = \sum_{n=-\infty}^{\lfloor x \rfloor} \mathbb{P}(X = n).$$

Suppose a and b are real numbers with $a \leq b$, and let $a_0 = \lfloor a \rfloor$ and $b_0 = \lfloor b \rfloor$. Then

$$F(b) - F(a) = \sum_{n=-\infty}^{b_0} \mathbb{P}(X = n) - \sum_{n=-\infty}^{a_0} \mathbb{P}(X = n) = \sum_{n=a_0+1}^{b_0} \mathbb{P}(X = n).$$

Since P is a probability distribution, we have $\mathbb{P}(X = n) \geq 0$ for all n, and since $a_0 \leq b_0$, $b_0 - (a_0 + 1) \geq -1$,[2] and so $\sum_{n=a_0+1}^{b_0} \mathbb{P}(X = n) \geq 0$. This completes the proof.

(b) $\lim_{b\to\infty} F(b) = 1$.

Solution: We have

$$\lim_{b\to\infty} F(b) = \sum_{n=-\infty}^{\infty} \mathbb{P}(X = n) = 1$$

by the axioms of probability.

(c) $\lim_{b\to-\infty} F(b) = 0$.

Solution: Let X_b denote the set of events less than or equal to b. That is,

$$X_b = (-\infty, b] \cap \mathbb{Z},$$

Then as $b \to -\infty$, $|X_b| \to 0$, meaning the number of terms contributing the the sum vanishes in the limit. Since $\mathbb{P}(X = x) \geq 0$ for each x in X_b, we have $\sum_{n=-\infty}^{b} \mathbb{P}(X = n) \to 0$ as $b \to -\infty$.

[2] Recall that $\sum_0^{-1} a_n = 0$ by natural convention (cf. *empty sum*).

(d) F is right continuous. That is, for any b and any decreasing sequence b_n, $n \geq 1$, that converges to b, $\lim_{n\to\infty} F(b_n) = F(b)$.

Solution: Let (b_n) be a decreasing sequence of real numbers with limit b and let $\{x_n\}_{n=-\infty}^{\infty}$ be the values the random variable takes on. There are two cases: either $x_i < b < x_{i+1}$ for some i, or $b = x_i$ for some i. If we are in the former case, then F is continuous on a neighborhood of b (in particular, if $x_i < \alpha < b < \beta < x_{i+1}$ then F is constant on $[\alpha, \beta]$), so it is indeed right-continuous.

Suppose $b = x_i$ for some i. Then F is constant on $[b, z]$ for any $b < z < x_{i+1}$. If $b_n \to b$ then for all $\varepsilon > 0$ there exists a natural number N such that $n > N$ implies $|b_n - b| = b_n - b < \varepsilon$. Choose $\varepsilon = z - b$. Then for all $n > N$ we have $b < b_n < \varepsilon + b < z$, and so $F(b_n) = F(b)$. Since $b_n \to b$, $F(b_n) \to F(b)$, which completes the proof.

Problem Set 5: Continuous Distributions

1. Suppose that X is an exponential random variable with parameter λ. Find the probability density function of $Y = \sqrt{X}$. What kind of random variable is Y?

   Solution: Let f_X denote the probability density function for X and f_Y for Y. Suppose $Y = g(X)$ for some monotone function g. I claim that

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|.$$

   Proof. Since g is monotone on its domain, g is a bijection and so g^{-1} exists. We have

$$\mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X \le g^{-1}y).$$

   If F_i is the cumulative distribution function for $i \in \{X, Y\}$, then the above shows $F_Y(y) = F_X(g^{-1}(Y))$. Since $F_i(z) = \int_{-\infty}^{z} f_i(\xi)\,\mathrm{d}\xi$, we differentiate both sides with respect to y to deduce

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}\right| = f_X(g^{-1}(y))\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|.$$

   The absolute value is included to ensure that $f_y(y) > 0$ in the event g^{-1} is monotonically decreasing. This completes the proof. $\square$

   In this case, $Y = g(X) = \sqrt{X}$ is monotonically increasing on $[0, \infty)$, and so

$$f_Y(y) = f_X(y^2)\left|\frac{\mathrm{d}(y^2)}{\mathrm{d}y}\right| = 2\lambda y e^{-\lambda y^2}.$$

   According the Wikipedia, this is a Rayleigh distribution.

2. If X is an exponential random variable with mean $1/\lambda$, show that $E(X^k) = k!/\lambda^k$, for $k = 1, 2, 3, \ldots$.

   Solution: Before conducting the proof, it will be helpful to prove the following fact:

$$\int_0^\infty x^k e^{-\lambda x}\,\mathrm{d}x = \frac{k!}{\lambda^{k+1}}.$$

   Proof. First we will determine a recursion relation for the integral, and then we will extrapolate. For $k > 1$, we integrate by parts (differentiating x^k and integrating $e^{-\lambda x}$) to get:

$$\int_0^\infty x^k e^{-\lambda x}\,\mathrm{d}x = -\frac{1}{\lambda}\,x^k e^{-\lambda x}\Big|_0^\infty + \frac{k}{\lambda}\int_0^\infty x^{k-1}e^{-\lambda x}\,\mathrm{d}x = \frac{k}{\lambda}\int_0^\infty x^{k-1}e^{-\lambda x}\,\mathrm{d}x\,.$$

   Hence, for arbitrarily large k, we find

$$\begin{aligned}\int_0^\infty x^k e^{-\lambda x}\,\mathrm{d}x &= \frac{k}{\lambda}\int_0^\infty x^{k-1}e^{-\lambda x}\,\mathrm{d}x = \frac{k(k-1)}{\lambda^2}\int_0^\infty x^{k-2}e^{-\lambda x}\,\mathrm{d}x \\ &= \cdots = \frac{k!}{\lambda^k}\int_0^\infty e^{-\lambda x}\,\mathrm{d}x = \frac{k!}{\lambda^k}\left[-\frac{1}{\lambda}e^{-\lambda x}\right]\Big|_{x=0}^\infty = \frac{k!}{\lambda^{k+1}}.\end{aligned}$$

   This completes the proof. $\square$

Let $X \sim \text{Exp}(\lambda)$. Then the probability density function is $f(x;\lambda) = \lambda e^{-\lambda x}$ for $x \geq 0$ and zero elsewhere. Using the law of the unconscious statistician, the expectation value is

$$\mathrm{E}(X^k) = \int_0^\infty x^k f(x;\lambda)\,\mathrm{d}x = \int_0^\infty x^k \cdot \lambda e^{-\lambda x}\,\mathrm{d}x = \lambda \cdot \frac{k!}{\lambda^{k+1}} = \frac{k!}{\lambda^k},$$

which is what we wanted to prove.

3. Let X be a random variable that takes on values between 0 and c. That is, $\mathbb{P}(0 \leq X \leq c) = 1$. Show that $\text{Var}(X) \leq c^2/4$.

Solution: We begin by proving the following claim:

$$\mathrm{E}(X^2) \leq c\mathrm{E}(X).$$

Proof. We integrate by parts, choosing to differentiate x and integrate $xf(x)$. We get

$$\mathrm{E}(X^2) = \int_0^c x^2 f(x)\,\mathrm{d}x = \left[x\int_0^x \xi f(\xi)\,\mathrm{d}\xi\right]_0^c - \int_0^c \int_0^x \xi f(\xi)\,\mathrm{d}\xi\,\mathrm{d}x$$
$$= c\int_0^c \xi f(\xi)\,\mathrm{d}\xi - K \leq c\int_0^c \xi f(\xi)\,\mathrm{d}\xi\,.$$

The last step follows from the fact that $xf(x) \geq 0$ in $[0,c]$ implies $K \geq 0$. The last integral is precisely $\mathrm{E}(X)$, so this completes the proof. □

Using the claim above, we find

$$\text{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) \leq c\mathrm{E}(X) - \mathrm{E}^2(X) = \mathrm{E}(X)[c - \mathrm{E}(X)].$$

This quantity is seen to be maximized[3] when $\mathrm{E}(X) = c/2$, and so $\text{Var}(X) \leq c^2/4$, as expected.

4. Properties of the normal distribution.

(a) Prove that the normal distribution expectation and variance is μ, σ^2.

Solution: If $X \sim \text{Normal}(\mu, \sigma^2)$, then the probability density function takes the form

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Before determining the expected value and variance, we need a few integral results. For $k > 0$, I claim the following:

$$\int_{-\infty}^{\infty} e^{-kx^2}\,\mathrm{d}x = \sqrt{\frac{\pi}{k}} \tag{1}$$
$$\int_{-\infty}^{\infty} xe^{-kx^2}\,\mathrm{d}x = 0 \tag{2}$$
$$\int_{-\infty}^{\infty} x^2 e^{-kx^2}\,\mathrm{d}x = -\sqrt{\frac{\pi}{4k^3}}. \tag{3}$$

[3] Since it is quadratic in $\mathrm{E}(X)$, one could just compute the vertex factorization and read off the maximal $\mathrm{E}(X)$, or one could take a derivative and set it equal to zero, verifying the maxima with the second derivative test. At this point, I'm just too lazy to type.

Proof (1). Let I denote the integral in question. We square I and convert to polar coordinates:

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-k(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\pi}\mathrm{d}\theta\int_0^{\infty} re^{-kr^2}\,\mathrm{d}r = -\frac{\pi}{k}\,e^{-kr^2}\Big|_{r=0}^{\infty} = \frac{\pi}{k}.$$

Hence $I = \sqrt{\pi/k}$. □

Proof (2). The integrand is odd and absolutely integrable, so the integral is zero. □

Proof (3). We use the result of integral (1) and Leibniz's integral rule. We have

$$\int_{-\infty}^{\infty} x^2e^{-kx^2}\,\mathrm{d}x = \int_{-\infty}^{\infty} -\frac{\partial}{\partial k}e^{-kx^2}\,\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}k}\int_{-\infty}^{\infty} e^{-kx^2}\,\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}k}\sqrt{\frac{\pi}{k}} = -\sqrt{\frac{\pi}{4k^3}},$$

which completes the proof. □

Now we can being the computation. The expected value $\mathrm{E}(X) = \int xf(x;\mu,\sigma^2)\,\mathrm{d}x$ is:

$$\begin{aligned}
\mathrm{E}(X) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} xe^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} (x+\mu)e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x + \frac{\mu}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}(0) + \frac{\mu}{\sqrt{2\pi\sigma^2}}\left(\sqrt{\frac{\pi}{1/2\sigma^2}}\right) \\
&= \mu,
\end{aligned}$$

as expected. Hence $\mathrm{E}^2(X) = \mu^2$, and $\mathrm{E}(X^2) = \int x^2f(x;\mu,\sigma^2)\,\mathrm{d}x$ is

$$\begin{aligned}
\mathrm{E}(X^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} x^2e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} (x+\mu)^2e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\left[\int_{-\infty}^{\infty} x^2e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x + 2\mu\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x + \mu^2\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x\right] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\left[\sqrt{\frac{\pi}{4/(2\sigma^2)^3}} + 2\mu(0) + \mu^2\left(\sqrt{\frac{\pi}{1/2\sigma^2}}\right)\right] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\left[\sqrt{2\pi\sigma^6} + \mu^2\sqrt{2\pi\sigma^2}\right] \\
&= \sigma^2 + \mu^2.
\end{aligned}$$

Hence $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = (\sigma^2+\mu^2) - \mu^2 = \sigma^2$, as expected.

(b) Prove that the standard normal distribution expectation and variance is 0, 1.

Solution: The standard normal has the following probability density function:

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

From part (a), we find $\mu = 0$ and $\sigma^2 = 1$.