Probability Distributions and Generating Functions

1 A Coherent View of Generating Functions

A generating function is a powerful mathematical tool that "encodes" an infinite sequence of numbers (e.g., a combinatorial sequence or a probability distribution) into a single function, usually a power series. The properties of this function (e.g., its derivatives, integrals, or algebraic form) reveal deep properties of the original sequence.

We can think of different generating functions as different "lenses" for viewing a sequence, each suited to a different purpose.

1.1 The Core Idea: From Sequence to Function

Let $\{a_n\}_{n=0}^{\infty}$ be a sequence of numbers. A generating function $G(x)$ is formed by:

$$G(x) = \sum_{n=0}^{\infty} a_n\, w_n(x)$$

where $w_n(x)$ is a set of "basis functions." The choice of $w_n(x)$ determines the type and properties of the generating function.

1.2 The Main Types of Generating Functions

1.2.1 Ordinary Generating Function (OGF)

- **Form:** $A(x) = \sum_{n=0}^{\infty} a_n x^n$
- **Sequence Element:** a_n
- **Basis:** $w_n(x) = x^n$
- **Purpose:** This is the workhorse of **unlabeled combinatorics** (counting). It is perfectly suited for problems involving "choosing" or "partitioning" items. The coefficient of x^n, denoted $[x^n]A(x)$, is the value a_n.
- **Key Property:** Multiplication corresponds to **convolution**. If $A(x)$ counts ways to build structure A and $B(x)$ counts ways to build structure B, then $C(x) = A(x)B(x)$ counts ways to build a combined structure C by $c_n = \sum_{k=0}^{n} a_k b_{n-k}$. This is the "sum rule" for disjoint parts: building a structure of size n by combining a part A of size k and a part B of size $n - k$.

1.2.2 Exponential Generating Function (EGF)

- **Form:** $E(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!}$
- **Sequence Element:** a_n
- **Basis:** $w_n(x) = \frac{x^n}{n!}$
- **Purpose:** This is the standard tool for **labeled combinatorics**. It is used for counting structures (like permutations, graphs, or derangements) where the underlying n elements are distinct. The coefficient of $x^n/n!$ is the value a_n.
- **Key Property:** Multiplication corresponds to **labeled products**. If $E_A(x)$ counts ways to build structure A on n labels and $E_B(x)$ counts ways to build structure B, then $E(x) = E_A(x)E_B(x)$ counts ways to form a new structure by:
  1. Splitting n labels into two sets of size k and $n - k$ (in $\binom{n}{k}$ ways).
  2. Building structure A on the k labels (in a_k ways).

3. Building structure B on the $n-k$ labels (in b_{n-k} ways).

This gives the convolution $c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}$, which is precisely what the product $E_A(x)E_B(x)$ encodes.

1.2.3 Probability Generating Function (PGF)

- **Form:** $G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} P(X=k)s^k$
- **Sequence Element:** $a_k = P(X=k)$, the probability mass function (PMF).
- **Basis:** $w_k(s) = s^k$
- **Purpose:** This is used *only* for non-negative, integer-valued random variables (RVs). It is, in fact, just an **OGF of the PMF**.
- **Key Property:**
    - $G_X(1) = \sum P(X=k) = 1$.
    - $\mathbb{E}[X] = G'_X(1)$ (First derivative at 1 gives the mean).
    - $\mathbb{E}[X(X-1)] = G''_X(1)$ (Second derivative at 1 gives the second factorial moment).
    - $\mathrm{Var}(X) = G''_X(1) + G'_X(1) - [G'_X(1)]^2$.
    - For independent X, Y, $G_{X+Y}(s) = \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X s^Y] = \mathbb{E}[s^X]\mathbb{E}[s^Y] = G_X(s)G_Y(s)$. The PGF of a sum is the product of the PGFs.

1.2.4 Moment Generating Function (MGF)

- **Form:** $M_X(t) = \mathbb{E}[e^{tX}]$
- **Sequence Element:** Not a simple sequence. It's an expectation.
- **Basis:** (Continuous case) $w_x(t) = e^{tx}$. $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$. This is the **Laplace Transform** of the PDF $f_X(x)$, with a sign flip in the exponent.
- **Purpose:** A powerful tool for *any* RV (discrete or continuous), provided the expectation exists in a neighborhood of $t=0$.
- **Key Property:**
    - $M_X(0) = \mathbb{E}[e^0] = 1$.
    - $M_X^{(n)}(0) = \frac{d^n}{dt^n}\mathbb{E}[e^{tX}]\Big|_{t=0} = \mathbb{E}\left[\frac{d^n}{dt^n}e^{tX}\right]\Big|_{t=0} = \mathbb{E}[X^n e^{tX}]\Big|_{t=0} = \mathbb{E}[X^n]$.
    - The n-th derivative at 0 gives the n-th **raw moment**. This is its namesake.
    - For independent X, Y, $M_{X+Y}(t) = M_X(t)M_Y(t)$.
    - (Uniqueness) If $M_X(t)$ exists in an open interval containing 0, it uniquely determines the distribution. This is crucial for proving theorems (like the CLT).

1.2.5 Characteristic Function (CF)

- **Form:** $\varphi_X(t) = \mathbb{E}[e^{itX}]$ (where $i = \sqrt{-1}$)
- **Basis:** (Continuous case) $w_x(t) = e^{itx}$. $\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,dx$. This is the **Fourier Transform** of the PDF $f_X(x)$.
- **Purpose:** This is the most general and robust generating function.

- **Key Property:**
  - It **always exists** for any RV X, because $|e^{itX}| = |\cos(tX)+i\sin(tX)| = \sqrt{\cos^2(tX)+\sin^2(tX)} = 1$. The expectation is of a bounded function, so it always converges. This is its main advantage over the MGF.
  - $\varphi_X^{(n)}(0) = i^n \mathbb{E}[X^n]$. It also generates moments.
  - For independent X, Y, $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.
  - (Uniqueness) The CF *always* uniquely determines the distribution.
  - (Lévy's Continuity Theorem) A sequence of RVs X_n converges in distribution to X if and only if their CFs converge pointwise: $\varphi_{X_n}(t) \to \varphi_X(t)$ for all t. This is the workhorse of modern probability theory.

1.2.6 Cumulant Generating Function (CGF)

- **Form:** $K_X(t) = \log M_X(t)$
- **Purpose:** A "helper" function derived from the MGF. It simplifies calculations involving sums of independent random variables.
- **Key Property:**
  - For independent X, Y, $M_{X+Y}(t) = M_X(t)M_Y(t)$.
  - Taking the log: $K_{X+Y}(t) = \log(M_{X+Y}(t)) = \log(M_X(t)M_Y(t)) = \log(M_X(t)) + \log(M_Y(t)) = K_X(t) + K_Y(t)$.
  - The CGF of a sum is the **sum of the CGFs**. Additivity is often simpler than multiplication.
  - The derivatives at 0 generate the **cumulants** κ_n.
  - $K_X^{(n)}(0) = \kappa_n$.
  - $\kappa_1 = K_X'(0) = \mathbb{E}[X]$ (Mean)
  - $\kappa_2 = K_X''(0) = \mathrm{Var}(X)$ (Variance)
  - $\kappa_3 = K_X'''(0) = \mathbb{E}[(X-\mu)^3]$ (Third central moment, related to Skewness)
  - $\kappa_4 = K_X^{(4)}(0) = \mathbb{E}[(X-\mu)^4] - 3(\mathrm{Var}(X))^2$ (Related to Kurtosis)

2 Generating Function Techniques: Solved Problems

This section provides solved problems for each type of generating function itself, distinct from their use in specific distributions.

2.1 Ordinary Generating Function (OGF)

Problem 2.1.1 (). Find the OGF for the sequence $a_n = c^n$ for some constant c.

Solution: By definition, the OGF is $A(x) = \sum_{n=0}^{\infty} a_n x^n$. Substituting $a_n = c^n$, we get:

$$A(x) = \sum_{n=0}^{\infty} c^n x^n = \sum_{n=0}^{\infty} (cx)^n$$

This is a standard geometric series with first term 1 and common ratio $r = cx$. The series converges to $\frac{1}{1-r}$ provided $|r| < 1$. Therefore, the OGF is:

$$A(x) = \frac{1}{1-cx}, \quad \text{for } |x| < 1/|c|$$

A special case is $a_n = 1$ (for $c = 1$), which gives the OGF $A(x) = \frac{1}{1-x} = 1 + x + x^2 + \dots$

Problem 2.1.2 (). Find the OGF for the Fibonacci sequence $\{F_n\}_{n=0}^{\infty}$, defined by $F_0 = 0, F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$.

Solution: Let $F(x) = \sum_{n=0}^{\infty} F_n x^n$ be the OGF. We will use the recurrence relation.

$$F(x) = F_0 + F_1 x + \sum_{n=2}^{\infty} F_n x^n$$

$$F(x) = 0 + 1x + \sum_{n=2}^{\infty} (F_{n-1} + F_{n-2}) x^n$$

$$F(x) = x + \sum_{n=2}^{\infty} F_{n-1} x^n + \sum_{n=2}^{\infty} F_{n-2} x^n$$

Now we "shift" the indices to match the form $\sum F_k x^k = F(x)$. For the first sum, let $k = n - 1$. When $n = 2, k = 1$.

$$\sum_{n=2}^{\infty} F_{n-1} x^n = \sum_{k=1}^{\infty} F_k x^{k+1} = x \sum_{k=1}^{\infty} F_k x^k = x(F(x) - F_0) = xF(x)$$

For the second sum, let $k = n - 2$. When $n = 2, k = 0$.

$$\sum_{n=2}^{\infty} F_{n-2} x^n = \sum_{k=0}^{\infty} F_k x^{k+2} = x^2 \sum_{k=0}^{\infty} F_k x^k = x^2 F(x)$$

Substitute these back into the equation for $F(x)$:

$$F(x) = x + xF(x) + x^2 F(x)$$

Now, we solve for $F(x)$:

$$F(x) - xF(x) - x^2 F(x) = x$$
$$F(x)(1 - x - x^2) = x$$
$$F(x) = \frac{x}{1 - x - x^2}$$

This is the closed-form OGF for the Fibonacci sequence. One could use partial fraction decomposition on this rational function to find Binet's formula for F_n.

Problem 2.1.3 (). Find the OGF for $p(n)$, the number of ways to partition an integer n (i.e., write n as a sum of positive integers, where order does not matter). For example, $p(4) = 5$ because $4 = 4$, $3 + 1$, $2 + 2$, $2 + 1 + 1$, $1 + 1 + 1 + 1$.

Solution: Let $P(x) = \sum_{n=0}^{\infty} p(n)x^n$ be the OGF, with $p(0) = 1$ (the empty partition). A partition of n can be described by how many 1's it uses, how many 2's, how many 3's, and so on. Let $n = c_1 \cdot 1 + c_2 \cdot 2 + c_3 \cdot 3 + \dots$, where $c_k \geq 0$. We can represent the choice of c_1 (how many 1's to use) by the OGF:

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

The term $x^{c_1 \cdot 1}$ in this factor "contributes" c_1 to the total sum n. Similarly, the choice of how many 2's to use is represented by:

$$(1 + x^2 + x^{2+2} + x^{2+2+2} + \dots) = (1 + x^2 + x^4 + x^6 + \dots) = \frac{1}{1 - x^2}$$

And for k's:

$$\frac{1}{1 - x^k}$$

Since the total n is the sum of these parts ($n = \sum c_k \cdot k$), the OGF for $p(n)$ is the *product* of the OGFs for each choice, as multiplication of OGFs corresponds to convolution (summing the exponents).

$$P(x) = \left(\frac{1}{1 - x}\right)\left(\frac{1}{1 - x^2}\right)\left(\frac{1}{1 - x^3}\right)\cdots$$

Therefore, the OGF for the partition function $p(n)$ is the infinite product:

$$P(x) = \prod_{k=1}^{\infty} \frac{1}{1 - x^k} = \prod_{k=1}^{\infty} (1 - x^k)^{-1}$$

This is a famous result by Euler. Extracting the coefficient $[x^n]P(x)$ is highly non-trivial but the OGF itself has a compact, elegant form.

2.2 Exponential Generating Function (EGF)

Problem 2.2.1 (). Find the EGF for the sequence $a_n = 1$ for all $n \geq 0$.

Solution: By definition, the EGF is $E(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!}$. Substituting $a_n = 1$, we get:

$$E(x) = \sum_{n=0}^{\infty} 1 \cdot \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

This is the Maclaurin series for the exponential function e^x. Therefore, the EGF is:

$$E(x) = e^x$$

This result is fundamental: e^x is the EGF for "choosing any number of labeled items and arranging them" (or in this case, just choosing them, as $a_n = 1$ is the number of permutations of n elements in a set of size n, which is $n!$, no, $a_n = 1$ is just the sequence of all ones). If $a_n = n!$ (permutations), the EGF is $\sum n! \frac{x^n}{n!} = \sum x^n = \frac{1}{1-x}$. If $a_n = 1$ (just "a single structure exists for n items"), the EGF is e^x.

Problem 2.2.2 (). Find the EGF for the Bell numbers B_n. B_n is the number of ways to partition a set of n labeled elements. For example, $B_3 = 5$ for the set $\{1, 2, 3\}$: $\{\{1, 2, 3\}\}$, $\{\{1, 2\}, \{3\}\}$, $\{\{1, 3\}, \{2\}\}$, $\{\{2, 3\}, \{1\}\}$, $\{\{1\}, \{2\}, \{3\}\}$.

Solution: Let $B(x) = \sum_{n=0}^{\infty} B_n \frac{x^n}{n!}$ be the EGF for Bell numbers, with $B_0 = 1$. A set partition of n elements is formed by placing the first element $\{1\}$ into a block, and then partitioning the remaining $n-1$ elements in one of B_{n-1} ways... this gets complicated. Let's use a structural approach. A set partition is a *set of non-empty sets* (the blocks). Let $A(x)$ be the EGF for a *single non-empty set.* A non-empty set of k labeled items can be formed in $a_k = 1$ way for $k \geq 1$, and $a_0 = 0$. The EGF for this is:

$$A(x) = \sum_{k=1}^{\infty} 1 \cdot \frac{x^k}{k!} = \left(\sum_{k=0}^{\infty} \frac{x^k}{k!} \right) - \frac{x^0}{0!} = e^x - 1$$

A partition of a set is a "set of these A structures." The "set of" construction for labeled objects corresponds to exp() of the EGF for the component. If $C(x)$ is the EGF for a component, $D(x) = \exp(C(x))$ is the EGF for a "set of C components." Here, the component is a non-empty block, with EGF $A(x) = e^x - 1$. Therefore, the EGF for Bell numbers (partitions, which are sets of blocks) is:

$$B(x) = \exp(A(x)) = \exp(e^x - 1)$$

This is a beautiful and compact result.

Problem 2.2.3 (). Find the EGF for D_n, the number of derangements of n elements (permutations with no fixed points).

Solution: Let $D(x) = \sum_{n=0}^{\infty} D_n \frac{x^n}{n!}$ be the EGF for derangements ($D_0 = 1$). A *permutation* (total $n!$ of them) can be decomposed into a *set of fixed points* and a *derangement* on the remaining points. Let $\Pi(x)$ be the EGF for all permutations, $a_n = n!$.

$$\Pi(x) = \sum_{n=0}^{\infty} n! \frac{x^n}{n!} = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

Let $F(x)$ be the EGF for "a set of fixed points." A fixed point is just a single element $k \to k$. The EGF for a *single* fixed point is just x (one choice of size 1). The EGF for a *set* of fixed points is $\exp(x)$. (Alternatively, $a_n = 1$ for "a set of n fixed points", so $F(x) = \sum_{n=0}^{\infty} 1 \cdot \frac{x^n}{n!} = e^x$). A permutation is a product of these two structures: (Set of Fixed Points) $\times$ (Derangement). By the EGF product rule for labeled structures:

$$\Pi(x) = F(x) \cdot D(x)$$

We can now solve for $D(x)$:

$$\frac{1}{1-x} = e^x \cdot D(x)$$

$$D(x) = \frac{e^{-x}}{1-x}$$

We can find D_n by taking the convolution: $e^{-x} = \sum_{n=0}^{\infty} (-1)^n \frac{x^n}{n!}$ and $\frac{1}{1-x} = \sum_{n=0}^{\infty} n! \frac{x^n}{n!}$ (using EGF form for $a_n = n!$). $D(x)$ is the product of $E_A(x) = \sum \frac{(-1)^n}{n!} x^n$ and $E_B(x) = \sum \frac{1}{n!} x^n$? No, that's e^{-x} and e^x. Let's use the OGF form for $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$. $D(x) = \left(\sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k!} \right) \left(\sum_{j=0}^{\infty} x^j \right)$ We want the coefficient of $x^n/n!$ in $D(x)$. $[x^n]D(x) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \cdot 1 = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$ $D_n = n![x^n]D(x)$? No, $D_n = n! \cdot (\text{coeff of } x^n)$. $D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$, which is the correct formula for derangements. The EGF is $D(x) = \frac{e^{-x}}{1-x}$.

2.3 Probability Generating Function (PGF)

Problem 2.3.1 (). Let $X \sim \text{Bernoulli}(p)$. Find its PGF $G_X(s)$ and use it to calculate $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution: A Bernoulli RV has $P(X = 1) = p$ and $P(X = 0) = 1 - p$. By definition, $G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} P(X = k) s^k$.

$$G_X(s) = P(X = 0)s^0 + P(X = 1)s^1 = (1 - p) + ps$$

This is the PGF. Now we find the derivatives at $s = 1$. First derivative:

$$G'_X(s) = \frac{d}{ds}(1 - p + ps) = p$$

$$\mathbb{E}[X] = G'_X(1) = p$$

Second derivative:

$$G''_X(s) = \frac{d}{ds}(p) = 0$$

$$\mathbb{E}[X(X-1)] = G''_X(1) = 0$$

Now, we find the variance:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

We know $\mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X] = 0$. So, $\mathbb{E}[X^2] = \mathbb{E}[X] = p$.

$$\text{Var}(X) = p - p^2 = p(1-p)$$

This confirms the known mean and variance.

Problem 2.3.2 (). Let $N \sim \text{Poisson}(\lambda)$ be the number of customers that arrive at a store. Each customer, independently, makes a purchase with probability p. Let $X_i \sim \text{Bernoulli}(p)$ be the purchase decision of the i-th customer. Let $K = \sum_{i=1}^{N} X_i$ be the total number of purchases. Find the PGF of K and identify its distribution.

Solution: This is a "compound distribution" problem. K is a random sum of random variables. We can find its PGF $G_K(s)$ using the law of total expectation, also known as Wald's Identity for PGFs:

$$G_K(s) = \mathbb{E}[s^K] = \mathbb{E}[\mathbb{E}[s^K|N]]$$

Given $N = n$, $K = \sum_{i=1}^{n} X_i$, which is a sum of n i.i.d. Bernoulli(p) variables. This is a Binomial(n, p) distribution. The PGF of a single $X_i \sim \text{Bern}(p)$ is $G_X(s) = (1 - p + ps)$. The PGF of $K|N = n \sim \text{Bin}(n, p)$ is $G_{K|N=n}(s) = (G_X(s))^n = (1 - p + ps)^n$. Now we substitute this back into the expectation over N:

$$G_K(s) = \mathbb{E}_N[(1 - p + ps)^N]$$

This has the form $\mathbb{E}[a^N]$ where $a = (1-p+ps)$. Recall the PGF of $N \sim \text{Poisson}(\lambda)$ is $G_N(s) = \mathbb{E}[s^N] = e^{\lambda(s-1)}$. So, $\mathbb{E}[a^N]$ is just $G_N(a)$:

$$G_K(s) = G_N(1 - p + ps) = e^{\lambda((1-p+ps)-1)}$$

$$G_K(s) = e^{\lambda(ps-p)} = e^{\lambda p(s-1)}$$

We recognize this as the PGF of a Poisson distribution with parameter λp. Therefore, $K \sim \text{Poisson}(\lambda p)$. This is known as the "Poisson thinning" property.

Problem 2.3.3 (). Consider a simple Galton-Watson branching process. Let $Z_0 = 1$ (one ancestor). Let X be the random variable for the number of offspring of any individual, with PGF $G_X(s) = \mathbb{E}[s^X]$. Let Z_n be the population size in generation n. Find the PGF of Z_n, $G_{Z_n}(s)$, in terms of $G_X(s)$.

Solution: Let $G_{Z_n}(s)$ be the PGF for the population size Z_n. $Z_0 = 1$, so $G_{Z_0}(s) = \mathbb{E}[s^{Z_0}] = s^1 = s$. Z_1 is the number of offspring of the single ancestor, so $Z_1 \sim X$.

$$G_{Z_1}(s) = G_X(s)$$

Now consider Z_2. Z_2 is the total number of offspring from the Z_1 individuals in the first generation. Let $X_i^{(1)}$ be the number of offspring of the i-th individual in generation 1. Then $Z_2 = \sum_{i=1}^{Z_1} X_i^{(1)}$. This is a random sum, just like in the previous (complex) problem. We can use the same composition property.

$$G_{Z_2}(s) = \mathbb{E}[s^{Z_2}] = \mathbb{E}[\mathbb{E}[s^{Z_2}|Z_1]]$$

Given $Z_1 = k$, $Z_2 = \sum_{i=1}^{k} X_i^{(1)}$. The PGF of this sum is $(G_X(s))^k$.

$$G_{Z_2}(s) = \mathbb{E}_{Z_1}[(G_X(s))^{Z_1}]$$

This has the form $\mathbb{E}[a^{Z_1}]$ where $a = G_X(s)$. By definition of $G_{Z_1}(s)$, this is $G_{Z_1}(G_X(s))$. Since $G_{Z_1}(s) = G_X(s)$, we have:

$$G_{Z_2}(s) = G_X(G_X(s)) = (G_X \circ G_X)(s)$$

We can see a pattern emerging. Let's assume $G_{Z_n}(s) = G_{Z_{n-1}}(G_X(s))$. Z_n is the sum of the offspring of the Z_{n-1} individuals from the previous generation: $Z_n = \sum_{i=1}^{Z_{n-1}} X_i^{(n-1)}$. Using the same logic (Wald's Identity for PGFs):

$$G_{Z_n}(s) = \mathbb{E}[s^{Z_n}] = \mathbb{E}_{Z_{n-1}}[(G_X(s))^{Z_{n-1}}]$$

This is exactly the definition of $G_{Z_{n-1}}(s)$ evaluated at the point $a = G_X(s)$.

$$G_{Z_n}(s) = G_{Z_{n-1}}(G_X(s))$$

This gives the recursive relation. $G_{Z_1}(s) = G_X(s)$ $G_{Z_2}(s) = G_X(G_X(s))$ $G_{Z_3}(s) = G_{Z_2}(G_X(s)) = G_X(G_X(G_X(s)))$ The PGF for the population in generation n is the n-fold composition of the offspring PGF with itself.

$$G_{Z_n}(s) = (G_X \circ \cdots \circ G_X)(s) \quad (n \text{ times})$$

2.4 Moment Generating Function (MGF)

Problem 2.4.1 (). Let $X \sim \text{Exponential}(\lambda)$, with PDF $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Find the MGF $M_X(t)$ and use it to find $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$.

Solution: By definition, $M_X(t) = \mathbb{E}[e^{tX}]$.

$$M_X(t) = \int_0^\infty e^{tx}(\lambda e^{-\lambda x})\,dx = \lambda \int_0^\infty e^{(t-\lambda)x}\,dx$$

This integral converges only if the exponent is negative, i.e., $t - \lambda < 0$, or $t < \lambda$. Assuming $t < \lambda$:

$$M_X(t) = \lambda \left[\frac{e^{(t-\lambda)x}}{t-\lambda}\right]_0^\infty = \lambda \left(\lim_{x\to\infty} \frac{e^{(t-\lambda)x}}{t-\lambda} - \frac{e^0}{t-\lambda}\right)$$

Since $t - \lambda < 0$, $\lim_{x\to\infty} e^{(t-\lambda)x} = 0$.

$$M_X(t) = \lambda\left(0 - \frac{1}{t-\lambda}\right) = \frac{-\lambda}{t-\lambda} = \frac{\lambda}{\lambda - t}$$

This is the MGF, valid for $t < \lambda$. Now we find moments by differentiating at $t = 0$:

$$M_X'(t) = \frac{d}{dt}\lambda(\lambda - t)^{-1} = \lambda(-1)(\lambda - t)^{-2}(-1) = \frac{\lambda}{(\lambda - t)^2}$$

$$\mathbb{E}[X] = M_X'(0) = \frac{\lambda}{(\lambda - 0)^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$M_X''(t) = \frac{d}{dt}\lambda(\lambda - t)^{-2} = \lambda(-2)(\lambda - t)^{-3}(-1) = \frac{2\lambda}{(\lambda - t)^3}$$

$$\mathbb{E}[X^2] = M_X''(0) = \frac{2\lambda}{(\lambda - 0)^3} = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$$

Problem 2.4.2 (). Let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ be independent. Prove that $Z = X + Y$ is also normally distributed and find its parameters, using MGFs.

Solution: The MGF for a general normal distribution $N(\mu, \sigma^2)$ is $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$. So, the MGFs for X and Y are:

$$M_X(t) = \exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right)$$
$$M_Y(t) = \exp\left(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right)$$

We want to find the MGF of the sum $Z = X + Y$. Because X and Y are independent, the MGF of the sum is the product of the MGFs:

$$M_Z(t) = M_{X+Y}(t) = M_X(t)M_Y(t)$$
$$M_Z(t) = \exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right) \cdot \exp\left(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right)$$

Combine the exponents:

$$M_Z(t) = \exp\left((\mu_1 t + \mu_2 t) + \left(\frac{1}{2}\sigma_1^2 t^2 + \frac{1}{2}\sigma_2^2 t^2\right)\right)$$
$$M_Z(t) = \exp\left((\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right)$$

We recognize this result. It is the MGF of a normal distribution with:

- Mean $\mu_Z = \mu_1 + \mu_2$
- Variance $\sigma_Z^2 = \sigma_1^2 + \sigma_2^2$

By the Uniqueness Property of MGFs, since $M_Z(t)$ has the form of a normal MGF, Z *must* be normally distributed. Therefore, $Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Problem 2.4.3 (). Prove the (weak) Law of Large Numbers (LLN) for a sequence of i.i.d. random variables $X_1, X_2, \ldots$ with mean $\mathbb{E}[X_i] = \mu$ and a finite MGF $M_X(t)$ in a neighborhood of 0.

Solution: The weak LLN states that the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ converges in probability to the true mean μ. That is, for any $\epsilon > 0$:

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

This can be proven using Chebyshev's inequality if we assume finite variance. However, a proof using MGFs is also possible and insightful (though it's more common for the CLT). Let's use a standard MGF-based proof using Markov's inequality, which is more direct. Let $M_X(t)$ be the common MGF, with $M_X'(0) = \mu$ and $M_X''(0) = \mathbb{E}[X^2]$. We assume $\sigma^2 = \mathbb{E}[X^2] - \mu^2 < \infty$. Let $S_n = \sum_{i=1}^n X_i$. The MGF of S_n is $M_{S_n}(t) = (M_X(t))^n$. The MGF of $\bar{X}_n = S_n/n$ is:

$$M_{\bar{X}_n}(t) = \mathbb{E}[e^{t(S_n/n)}] = \mathbb{E}[e^{(t/n)S_n}] = M_{S_n}(t/n) = (M_X(t/n))^n$$

As $n \to \infty$, $t/n \to 0$. We can use a Taylor expansion for $M_X(u)$ around $u = 0$:

$$M_X(u) = 1 + M_X'(0)u + O(u^2) = 1 + \mu u + O(u^2)$$

Let $u = t/n$:

$$M_X(t/n) = 1 + \mu(t/n) + O(1/n^2)$$

Now substitute this back into the MGF for $\bar{X}_n$:

$$M_{\bar{X}_n}(t) = \left(1 + \frac{\mu t}{n} + O(1/n^2)\right)^n$$

As $n \to \infty$, this converges to $\exp(\mu t)$.

$$\lim_{n\to\infty} M_{\bar{X}_n}(t) = \lim_{n\to\infty}\left(1 + \frac{\mu t}{n}\right)^n = e^{\mu t}$$

The function $M_Z(t) = e^{\mu t}$ is the MGF of a constant random variable $Z = \mu$ (a "degenerate" distribution at μ). By the MGF Uniqueness and Continuity Theorem, if the MGFs converge ($M_{\bar{X}_n}(t) \to M_Z(t)$), then the random variables converge in distribution ($\bar{X}_n \xrightarrow{d} Z$). Convergence in distribution to a constant is equivalent to convergence in probability to that constant. Therefore, $\bar{X}_n \xrightarrow{p} \mu$.

2.5 Characteristic Function (CF)

Problem 2.5.1 (). Find the CF of the standard Cauchy distribution, which has PDF $f(x) = \frac{1}{\pi(1+x^2)}$. Note that its MGF does not exist.

Solution: First, let's show the MGF $M_X(t) = \mathbb{E}[e^{tX}]$ does not exist for $t \neq 0$.

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\pi(1+x^2)} \, dx$$

For $t > 0$, as $x \to \infty$, e^{tx} grows exponentially while $1/(1+x^2)$ only decays polynomially. The e^{tx} term dominates, and the integral diverges. Similarly for $t < 0$ as $x \to -\infty$. Thus, MGF does not exist in any neighborhood of 0. Now, let's find the CF $\varphi_X(t) = \mathbb{E}[e^{itX}]$.

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi(1+x^2)} \, dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos(tx) + i\sin(tx)}{1+x^2} \, dx$$

Since $\sin(tx)/(1+x^2)$ is an odd function, its integral from $-\infty$ to ∞ is 0.

$$\varphi_X(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos(tx)}{1+x^2} \, dx$$

Since the integrand is even, $\int_{-\infty}^{\infty} = 2\int_0^{\infty}$.

$$\varphi_X(t) = \frac{2}{\pi} \int_{0}^{\infty} \frac{\cos(tx)}{1+x^2} \, dx$$

This is a standard, but difficult, integral, often solved using contour integration in complex analysis. The result is known as $\int_0^{\infty} \frac{\cos(ax)}{1+x^2} dx = \frac{\pi}{2} e^{-|a|}$. Assuming this result, we have $a = t$:

$$\varphi_X(t) = \frac{2}{\pi} \left(\frac{\pi}{2} e^{-|t|} \right) = e^{-|t|}$$

This CF exists for all t, demonstrating the robustness of the CF.

Problem 2.5.2 (). Let $X_1, \ldots, X_n$ be i.i.d. from a Cauchy$(0,1)$ distribution. Find the distribution of the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. What does this imply about the Law of Large Numbers?

Solution: We will use Characteristic Functions. From the previous problem, the CF for a single X_i is $\varphi_X(t) = e^{-|t|}$. Let $S_n = \sum_{i=1}^n X_i$. Since the X_i are independent, the CF of the sum is the product of the CFs:

$$\varphi_{S_n}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t) = (e^{-|t|})^n = e^{-n|t|}$$

Now we find the CF of the sample mean $\bar{X}_n = S_n/n$. We use the scaling property of CFs: $\varphi_{aY}(t) = \varphi_Y(at)$.

$$\varphi_{\bar{X}_n}(t) = \varphi_{S_n/n}(t) = \varphi_{S_n}(t/n)$$

Substitute t/n into the CF for S_n:

$$\varphi_{\bar{X}_n}(t) = e^{-n|t/n|} = e^{-n \cdot \frac{|t|}{n}} = e^{-|t|}$$

This result is remarkable. The characteristic function of the sample mean $\bar{X}_n$ is $e^{-|t|}$, which is *exactly the same* as the CF of a single observation X_1. By the Uniqueness Property of CFs, this means that the distribution of $\bar{X}_n$ is identical to the distribution of X_1.

$$\bar{X}_n \sim \text{Cauchy}(0,1)$$

This implies that averaging Cauchy variables does not reduce their variability or change their distribution at all. The sample mean does not converge to a single point (the "mean", which is undefined). This distribution *violates* the conditions for the Law of Large Numbers (it does not have a finite mean). The sample mean $\bar{X}_n$ does not converge in probability to any constant.

Problem 2.5.3 (). State and prove the Central Limit Theorem (CLT) using Lévy's Continuity Theorem and Characteristic Functions.

Solution: **Theorem (CLT):** Let $X_1, X_2, \dots$ be i.i.d. random variables with finite mean $\mathbb{E}[X_i] = \mu$ and finite variance $\mathrm{Var}(X_i) = \sigma^2 > 0$. Let $S_n = \sum_{i=1}^n X_i$ and $\bar{X}_n = S_n/n$. Then the standardized sample mean Z_n converges in distribution to a standard normal $N(0, 1)$:

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0, 1)$$

Proof using CFs: By Lévy's Continuity Theorem, it suffices to show that the CF of Z_n, $\varphi_{Z_n}(t)$, converges pointwise to the CF of $Z \sim N(0, 1)$, which is $\varphi_Z(t) = e^{-t^2/2}$. Let $Y_i = (X_i - \mu)/\sigma$. The Y_i are i.i.d. with $\mathbb{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) = 1$.

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sum(X_i - \mu)}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$$

Let $\varphi_Y(t)$ be the CF of a single Y_i. By independence, the CF of Z_n is:

$$\varphi_{Z_n}(t) = \varphi_{\frac{1}{\sqrt{n}}\sum Y_i}(t) = \varphi_{\sum Y_i}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Now we perform a Taylor expansion of $\varphi_Y(u)$ around $u = 0$. $\varphi_Y(u) = \mathbb{E}[e^{iuY}]$ $\varphi_Y'(u) = \mathbb{E}[iYe^{iuY}] \implies \varphi_Y'(0) = \mathbb{E}[iY] = i\mathbb{E}[Y] = 0$ $\varphi_Y''(u) = \mathbb{E}[i^2Y^2e^{iuY}] \implies \varphi_Y''(0) = \mathbb{E}[-Y^2] = -\mathbb{E}[Y^2] = -\mathrm{Var}(Y) = -1$ The Taylor expansion to second order is:

$$\varphi_Y(u) = \varphi_Y(0) + \varphi_Y'(0)u + \frac{\varphi_Y''(0)}{2!}u^2 + o(u^2)$$

$$\varphi_Y(u) = 1 + 0 \cdot u + \frac{-1}{2}u^2 + o(u^2) = 1 - \frac{u^2}{2} + o(u^2)$$

Now, let $u = t/\sqrt{n}$. As $n \to \infty$, $u \to 0$.

$$\varphi_Y\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right) = 1 - \frac{t^2}{2n} + o(1/n)$$

Substitute this back into the expression for $\varphi_{Z_n}(t)$:

$$\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + o(1/n)\right)^n$$

This is of the form $\lim_{n\to\infty}(1 + x/n)^n = e^x$. Here, $x = -t^2/2$.

$$\lim_{n\to\infty} \varphi_{Z_n}(t) = \exp\left(-\frac{t^2}{2}\right)$$

This is the characteristic function of a $N(0, 1)$ distribution. By Lévy's Continuity Theorem, $Z_n \xrightarrow{d} N(0, 1)$. This completes the proof.

---

3 Discrete Distributions

We now present standard discrete distributions, each with definitions, properties, and a set of solved problems.

3.1 Bernoulli Distribution

Definition and Properties

A random variable X follows a Bernoulli distribution with parameter p ($X \sim \text{Bernoulli}(p)$) if it represents a single trial with two outcomes (e.g., success/failure).

- **PMF:** $P(X = x) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$
- **PGF:** $G_X(s) = 1 - p + ps$
- **MGF:** $M_X(t) = 1 - p + pe^t$
- **CF:** $\varphi_X(t) = 1 - p + pe^{it}$
- **Mean:** $\mathbb{E}[X] = p$
- **Variance:** $\text{Var}(X) = p(1-p)$

Discussion: The Bernoulli is the fundamental building block for many other discrete distributions (Binomial, Geometric, Negative Binomial). It models any single yes/no event.

Solved Problems

Problem 3.1.1 (). A fair coin is tossed. Let $X = 1$ if heads, $X = 0$ if tails. What is $P(X = 1)$?

Solution: This is a Bernoulli trial with $p =$ probability of heads. For a fair coin, $p = 0.5$. $P(X = 1) = p = 0.5$.

Problem 3.1.2 (). Let $X \sim \text{Bernoulli}(p)$. Find the MGF of the random variable $Y = aX + b$, where a and b are constants.

Solution: We want to find $M_Y(t) = \mathbb{E}[e^{tY}]$.

$$M_Y(t) = \mathbb{E}[e^{t(aX+b)}] = \mathbb{E}[e^{atX+bt}] = \mathbb{E}[e^{atX} \cdot e^{bt}]$$

Since e^{bt} is a constant, we can pull it out of the expectation:

$$M_Y(t) = e^{bt}\mathbb{E}[e^{(at)X}]$$

The term $\mathbb{E}[e^{(at)X}]$ is just the MGF of X, $M_X(t)$, evaluated at the point (at). The MGF of $X \sim \text{Bernoulli}(p)$ is $M_X(t) = 1 - p + pe^t$. Therefore, $M_X(at) = 1 - p + pe^{(at)}$. Substituting this back, we get:

$$M_Y(t) = e^{bt}(1 - p + pe^{at})$$

Problem 3.1.3 (). Show that the Bernoulli(p) distribution is a member of the one-parameter exponential family.

Solution: A distribution is in the exponential family if its PMF (or PDF) $f(x|\theta)$ can be written in the form:

$$f(x|\theta) = h(x)\exp(\eta(\theta)T(x) - A(\theta))$$

Here, $x \in \{0,1\}$ and the parameter is $\theta = p$. The PMF is $f(x|p) = p^x(1-p)^{1-x}$. We want to manipulate this into the required form.

$$f(x|p) = \exp\left(\log(p^x(1-p)^{1-x})\right)$$
$$f(x|p) = \exp\left(x\log(p) + (1-x)\log(1-p)\right)$$
$$f(x|p) = \exp\left(x\log(p) + \log(1-p) - x\log(1-p)\right)$$
$$f(x|p) = \exp\left(x(\log(p) - \log(1-p)) + \log(1-p)\right)$$
$$f(x|p) = \exp\left(x\log\left(\frac{p}{1-p}\right) + \log(1-p)\right)$$

Now we identify the components:

- $h(x) = 1$ (since x only takes values 0 or 1, this is a constant, or can be absorbed)
- $\eta(p) = \log\left(\frac{p}{1-p}\right)$ (This is the "natural parameter," also known as the log-odds or logit).
- $T(x) = x$ (This is the "sufficient statistic").
- $A(p) = -\log(1-p) = \log\left(\frac{1}{1-p}\right)$.

We can express $A(p)$ in terms of η. Let $\eta = \log(p/(1-p))$. Then $e^\eta = p/(1-p) \implies e^\eta(1-p) = p \implies e^\eta - e^\eta p = p \implies e^\eta = p(1+e^\eta) \implies p = e^\eta/(1+e^\eta)$. And $1-p = 1-\frac{e^\eta}{1+e^\eta} = \frac{1}{1+e^\eta}$. So $A(\eta) = -\log\left(\frac{1}{1+e^\eta}\right) = \log(1+e^\eta)$. The canonical form is:

$$f(x|\eta) = 1 \cdot \exp(\eta \cdot x - \log(1+e^\eta))$$

Since the PMF can be written in this form, the Bernoulli(p) distribution is a member of the exponential family.

3.2 Binomial Distribution

Definition and Properties

A random variable $X \sim \text{Binomial}(n,p)$ represents the total number of successes in n independent Bernoulli(p) trials.

- **PMF:** $P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$ for $k \in \{0,1,\dots,n\}$
- **PGF:** $G_X(s) = (1-p+ps)^n$
- **MGF:** $M_X(t) = (1-p+pe^t)^n$
- **CF:** $\varphi_X(t) = (1-p+pe^{it})^n$
- **Mean:** $\mathbb{E}[X] = np$
- **Variance:** $\text{Var}(X) = np(1-p)$

Discussion: This is the sum of n i.i.d. Bernoulli(p) variables. The PGF/MGF/CF forms are n-th powers of the Bernoulli GFs, reflecting this sum.

Solved Problems

Problem 3.2.1 (). A student guesses randomly on a 10-question multiple-choice test. Each question has 4 options. What is the probability the student gets exactly 3 questions right?

Solution: This is a Binomial(n, p) problem. Number of trials (questions): $n = 10$. Probability of success (guessing correctly): $p = 1/4 = 0.25$. Number of successes we want: $k = 3$. Let X be the number of correct answers. $X \sim \text{Binomial}(10, 0.25)$. We use the PMF:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(X = 3) = \binom{10}{3} (0.25)^3 (1 - 0.25)^{10-3}$$

$$P(X = 3) = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} (0.25)^3 (0.75)^7$$

$$P(X = 3) = 120 \cdot (0.015625) \cdot (0.13348) \approx 0.2503$$

There is approximately a 25.03% chance of guessing exactly 3 questions right.

Problem 3.2.2 (). Use the MGF to prove that if $X \sim \text{Binomial}(n_1, p)$ and $Y \sim \text{Binomial}(n_2, p)$ are independent, then $Z = X + Y \sim \text{Binomial}(n_1 + n_2, p)$.

Solution: This is the additive property of the Binomial distribution. We use MGFs. The MGF for X is $M_X(t) = (1 - p + pe^t)^{n_1}$. The MGF for Y is $M_Y(t) = (1 - p + pe^t)^{n_2}$. Since X and Y are independent, the MGF of their sum $Z = X + Y$ is the product of their MGFs:

$$M_Z(t) = M_X(t) \cdot M_Y(t)$$

$$M_Z(t) = (1 - p + pe^t)^{n_1} \cdot (1 - p + pe^t)^{n_2}$$

Using rules of exponents, we add the powers:

$$M_Z(t) = (1 - p + pe^t)^{n_1 + n_2}$$

We immediately recognize this as the MGF of a Binomial distribution with parameters $n = n_1 + n_2$ and p. By the Uniqueness Property of MGFs, since $M_Z(t)$ has this form, Z must follow this distribution. Therefore, $Z \sim \text{Binomial}(n_1 + n_2, p)$.

Problem 3.2.3 (). Derive the Poisson(λ) distribution as a limit of the Binomial(n, p) distribution, using Probability Generating Functions (PGFs).

Solution: We are looking for the "Poisson limit" or "law of rare events." This occurs when $n \to \infty$ and $p \to 0$ simultaneously, such that the mean $\lambda = np$ remains constant. This implies $p = \lambda/n$. Let $X_n \sim \text{Binomial}(n, \lambda/n)$. Its PGF is:

$$G_{X_n}(s) = (1 - p + ps)^n = \left(1 - \frac{\lambda}{n} + s\frac{\lambda}{n}\right)^n$$

$$G_{X_n}(s) = \left(1 + \frac{\lambda(s-1)}{n}\right)^n$$

Now, we take the limit as $n \to \infty$.

$$\lim_{n\to\infty} G_{X_n}(s) = \lim_{n\to\infty} \left(1 + \frac{\lambda(s-1)}{n}\right)^n$$

This is a standard limit of the form $\lim_{n\to\infty}(1 + x/n)^n = e^x$. Here, $x = \lambda(s - 1)$.

$$\lim_{n\to\infty} G_{X_n}(s) = \exp(\lambda(s-1))$$

This resulting function, $G_Y(s) = e^{\lambda(s-1)}$, is the PGF of the Poisson(λ) distribution. By the PGF convergence theorem (a corollary of Lévy's Continuity Theorem), since the PGF of X_n converges to the PGF of $Y \sim$ Poisson(λ), the distribution X_n converges in distribution to Y.

$$\text{Binomial}(n, \lambda/n) \xrightarrow{d} \text{Poisson}(\lambda) \quad \text{as } n \to \infty$$

3.3 Multinomial Distribution

Definition and Properties

The Multinomial$(n, \mathbf{p})$ distribution generalizes the Binomial to r possible outcomes. It models n independent trials, where each trial can result in one of r categories with probabilities $\mathbf{p} = (p_1, p_2, \dots, p_r)$, where $\sum p_i = 1$. The RV $\mathbf{X} = (X_1, \dots, X_r)$ counts the number of times each category occurred.

- **PMF:** $P(X_1 = k_1, \dots, X_r = k_r) = \frac{n!}{k_1!k_2!\dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$, where $\sum k_i = n$.
- **MGF:** $M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[e^{\mathbf{t}\cdot\mathbf{X}}] = \left(\sum_{i=1}^r p_i e^{t_i}\right)^n$
- **CF:** $\varphi_{\mathbf{X}}(\mathbf{t}) = \left(\sum_{i=1}^r p_i e^{it_i}\right)^n$
- **Mean:** $\mathbb{E}[X_i] = np_i$ (for the i-th component)
- **Variance:** $\mathrm{Var}(X_i) = np_i(1 - p_i)$
- **Covariance:** $\mathrm{Cov}(X_i, X_j) = -np_ip_j$ for $i \neq j$.

Discussion: The marginal distribution of any single component X_i is Binomial(n, p_i). The negative covariance makes sense: if X_i is large (more outcomes in category i), then X_j must be smaller, since their sum is fixed at n.

Solved Problems

Problem 3.3.1 (). A fair six-sided die is rolled 12 times. What is the probability of rolling exactly two 1s, three 2s, and the other seven rolls being "3 or greater"?

Solution: This is a Multinomial problem with $n = 12$ trials. We can define $r = 3$ categories:

- Category 1: "Roll is 1". $p_1 = 1/6$. We want $k_1 = 2$.
- Category 2: "Roll is 2". $p_2 = 1/6$. We want $k_2 = 3$.
- Category 3: "Roll is 3, 4, 5, or 6". $p_3 = 4/6 = 2/3$. We must have $k_3 = 12 - 2 - 3 = 7$.

Check: $p_1 + p_2 + p_3 = 1/6 + 1/6 + 4/6 = 1$. $\sum k_i = 2 + 3 + 7 = 12 = n$. Now, we use the PMF:

$$P(X_1 = 2, X_2 = 3, X_3 = 7) = \frac{12!}{2!3!7!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^3\left(\frac{4}{6}\right)^7$$

$$P = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{2 \cdot (3 \cdot 2 \cdot 1)}\left(\frac{1}{6}\right)^5\left(\frac{4}{6}\right)^7$$

$$P = (11 \cdot 10 \cdot 9 \cdot 8)\left(\frac{1}{6^5}\right)\left(\frac{4^7}{6^7}\right) = 7920\frac{4^7}{6^{12}} = 7920\frac{16384}{2176782336} \approx 0.0000596$$

This is a very small probability.

Problem 3.3.2 (). Let $\mathbf{X} \sim$ Multinomial$(n, \mathbf{p})$ with r categories. Find the marginal distribution of X_1.

Solution: We can find the marginal distribution by "summing out" (or rather, "grouping") all other categories. Let's define a new trial with two outcomes:

- Success: Outcome is in category 1. Probability p_1.
- Failure: Outcome is *not* in category 1 (i.e., in categories 2, 3, ..., r). Probability $p_2 + p_3 + \dots + p_r = 1 - p_1$.

Now, X_1 is just the count of successes in n trials. This is a classic Binomial setup. The number of successes X_1 follows a Binomial(n, p_1) distribution. We can also prove this using the MGF. The MGF of $\mathbf{X}$ is $M_{\mathbf{X}}(t_1, \dots, t_r) = \left(\sum_{i=1}^r p_i e^{t_i}\right)^n$. The MGF of the marginal X_1 is found by setting all other $t_j = 0$ $(j \neq 1)$.

$$M_{X_1}(t_1) = M_{\mathbf{X}}(t_1, 0, 0, \dots, 0)$$

$$M_{X_1}(t_1) = \left(p_1 e^{t_1} + p_2 e^0 + p_3 e^0 + \cdots + p_r e^0\right)^n$$

$$M_{X_1}(t_1) = \left(p_1 e^{t_1} + (p_2 + p_3 + \cdots + p_r)\right)^n$$

Since $\sum p_i = 1$, we have $(p_2 + \cdots + p_r) = 1 - p_1$.

$$M_{X_1}(t_1) = \left(p_1 e^{t_1} + (1 - p_1)\right)^n$$

This is precisely the MGF of a Binomial(n, p_1) distribution. By the Uniqueness Property, $X_1 \sim$ Binomial(n, p_1).

Problem 3.3.3 (). Let $\mathbf{X} \sim$ Multinomial$(n, \mathbf{p})$ with $r = 3$. So $\mathbf{X} = (X_1, X_2, X_3)$. Find the conditional distribution of X_1 given $X_3 = k_3$.

Solution: We want to find $P(X_1 = k_1 | X_3 = k_3)$. By definition of conditional probability:

$$P(X_1 = k_1 | X_3 = k_3) = \frac{P(X_1 = k_1, X_3 = k_3)}{P(X_3 = k_3)}$$

First, let's find the numerator. $P(X_1 = k_1, X_3 = k_3)$ means X_2 must be $k_2 = n - k_1 - k_3$.

$$P(X_1 = k_1, X_2 = n - k_1 - k_3, X_3 = k_3) = \frac{n!}{k_1!(n - k_1 - k_3)!k_3!} p_1^{k_1} p_2^{n-k_1-k_3} p_3^{k_3}$$

This is the joint probability $P(X_1 = k_1, X_3 = k_3)$. Next, the denominator. As shown in the previous problem, the marginal X_3 is Binomial(n, p_3).

$$P(X_3 = k_3) = \binom{n}{k_3} p_3^{k_3} (1 - p_3)^{n-k_3} = \frac{n!}{k_3!(n - k_3)!} p_3^{k_3} (p_1 + p_2)^{n-k_3}$$

Now we divide the two. This is going to be algebraically intensive.

$$P(X_1 = k_1 | X_3 = k_3) = \frac{\frac{n!}{k_1!(n-k_1-k_3)!k_3!} p_1^{k_1} p_2^{n-k_1-k_3} p_3^{k_3}}{\frac{n!}{k_3!(n-k_3)!} p_3^{k_3} (p_1 + p_2)^{n-k_3}}$$

Many terms cancel. $n!$ and $k_3!$ cancel. $p_3^{k_3}$ cancels.

$$= \frac{(n - k_3)!}{k_1!(n - k_1 - k_3)!} \cdot \frac{p_1^{k_1} p_2^{n-k_1-k_3}}{(p_1 + p_2)^{n-k_3}}$$

Let $m = n - k_3$. This is the "remaining" number of trials after we *know* k_3 of them were category 3. The denominator exponent is $n - k_3 = (n - k_1 - k_3) + k_1$.

$$= \frac{m!}{k_1!(m - k_1)!} \cdot \frac{p_1^{k_1} p_2^{m-k_1}}{(p_1 + p_2)^m}$$

$$= \binom{m}{k_1} \frac{p_1^{k_1} p_2^{m-k_1}}{(p_1 + p_2)^m} = \binom{m}{k_1} \left(\frac{p_1}{p_1 + p_2}\right)^{k_1} \left(\frac{p_2}{p_1 + p_2}\right)^{m-k_1}$$

Let $p_1^* = \frac{p_1}{p_1+p_2}$. This is the *conditional probability* of being in category 1, given you are *not* in category 3. Let $p_2^* = \frac{p_2}{p_1+p_2} = 1 - p_1^*$. The conditional PMF is:

$$\binom{m}{k_1} (p_1^*)^{k_1} (1 - p_1^*)^{m-k_1}$$

This is the PMF of a Binomial(m, p_1^*) distribution. Therefore, $(X_1 | X_3 = k_3) \sim$ Binomial$\left(n - k_3, \frac{p_1}{p_1+p_2}\right)$. This makes intuitive sense: given k_3 outcomes are fixed, we are left with $n - k_3$ trials that must have been either category 1 or 2. The (rescaled) probability of being category 1 is $p_1/(p_1 + p_2)$.

3.4 Geometric Distribution

Definition and Properties

(Parameterization 1: X = number of trials until the first success).

- **PMF:** $P(X = k) = (1-p)^{k-1}p$ for $k \in \{1, 2, 3, \dots\}$
- **PGF:** $G_X(s) = \frac{ps}{1-s(1-p)}$
- **MGF:** $M_X(t) = \frac{pe^t}{1-e^t(1-p)}$
- **CF:** $\varphi_X(t) = \frac{pe^{it}}{1-e^{it}(1-p)}$
- **Mean:** $\mathbb{E}[X] = 1/p$
- **Variance:** $\text{Var}(X) = (1-p)/p^2$

Discussion: This is the "waiting time" distribution. It is the only discrete distribution with the **memoryless property**: $P(X > k+j|X > k) = P(X > j)$. The fact that you have already waited k trials does not change the probability distribution of waiting j more.

Solved Problems

Problem 3.4.1 (). A basketball player has a 70% free throw success rate ($p = 0.7$). What is the probability that they make their first free throw on their 3rd attempt?

Solution: This is a Geometric(0.7) distribution. We want $P(X = 3)$. This means 2 failures (F) followed by 1 success (S). F: $P(\text{Fail}) = 1 - p = 0.3$ S: $P(\text{Success}) = p = 0.7$

$$P(X = 3) = (1-p)^{3-1}p = (0.3)^2(0.7) = 0.09 \cdot 0.7 = 0.063$$

There is a 6.3% chance of this happening.

Problem 3.4.2 (). Prove the memoryless property of the Geometric distribution. That is, show $P(X > k+j|X > k) = P(X > j)$ for $j, k \geq 1$.

Solution: First, we need the "survival function" $P(X > k)$. $P(X > k)$ is the probability that the first k trials are all failures. The probability of one failure is $(1-p)$. The probability of k independent failures is $(1-p)^k$. So, $P(X > k) = (1-p)^k$. Now, let's analyze the conditional probability:

$$P(X > k+j|X > k) = \frac{P(X > k+j \text{ and } X > k)}{P(X > k)}$$

The event "X ¿ k+j" (e.g., waiting more than 10 trials) is a *subset* of the event "X ¿ k" (e.g., waiting more than 5 trials). If you waited more than 10 trials, you *definitely* waited more than 5. Therefore, the intersection "$(X > k+j)$ and $(X > k)$" is just "$(X > k+j)$".

$$P(X > k+j|X > k) = \frac{P(X > k+j)}{P(X > k)}$$

Now we use our survival function:

$$P(X > k+j|X > k) = \frac{(1-p)^{k+j}}{(1-p)^k} = \frac{(1-p)^k(1-p)^j}{(1-p)^k} = (1-p)^j$$

We also know that $P(X > j) = (1-p)^j$. Therefore, $P(X > k+j|X > k) = P(X > j)$. This proves the property. It means if you have already waited k trials for a success, the probability distribution for the *remaining* waiting time is identical to the original distribution. The process "forgets" how long it has been waiting.

Problem 3.4.3 (). Let $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(p)$ be independent. Find the distribution of $Z = X + Y$.

Solution: We can solve this using PGFs. X and Y are i.i.d., so $G_X(s) = G_Y(s) = \frac{ps}{1-s(1-p)}$. Since they are independent, the PGF of the sum $Z = X + Y$ is the product of the PGFs:

$$G_Z(s) = G_X(s)G_Y(s) = \left(\frac{ps}{1-s(1-p)}\right)^2 = \left(\frac{ps}{1-(1-p)s}\right)^2$$

We recognize this PGF. The PGF for a Negative Binomial(r, p) distribution (counting trials) is $G(s) = \left(\frac{ps}{1-(1-p)s}\right)^r$. Our $G_Z(s)$ matches this form with $r = 2$. Therefore, $Z = X + Y \sim \text{Negative Binomial}(r = 2, p)$. This makes intuitive sense: The Negative Binomial(r, p) distribution counts the total number of trials to get r successes. X is the number of trials to get the 1st success. Y is the number of trials to get the 2nd success *after* the 1st. By the memoryless property, this Y has the same distribution as X. $Z = X + Y$ is the total number of trials to get the 1st success, plus the number of additional trials to get the 2nd success. This is, by definition, the total number of trials to get 2 successes.

3.5 Negative Binomial Distribution

Definition and Properties

(Parameterization 1: X = number of trials to achieve r successes).

- **PMF:** $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$ for $k \in \{r, r+1, \dots\}$
- **PGF:** $G_X(s) = \left(\frac{ps}{1-s(1-p)}\right)^r$
- **MGF:** $M_X(t) = \left(\frac{pe^t}{1-e^t(1-p)}\right)^r$
- **CF:** $\varphi_X(t) = \left(\frac{pe^{it}}{1-e^{it}(1-p)}\right)^r$
- **Mean:** $\mathbb{E}[X] = r/p$
- **Variance:** $\text{Var}(X) = r(1-p)/p^2$

Discussion: This is the sum of r independent Geometric(p) variables. When $r = 1$, it simplifies to the Geometric distribution.

Solved Problems

Problem 3.5.1 (). The basketball player from before ($p = 0.7$) needs to make $r = 5$ free throws to win a prize. What is the probability that it takes *exactly* $k = 6$ attempts?

Solution: This is a Negative Binomial$(r = 5, p = 0.7)$ problem. We want $P(X = 6)$. This means the 6th attempt must be a success, and in the first $k - 1 = 5$ attempts, there must have been $r - 1 = 4$ successes. We can use the PMF:

$$P(X = 6) = \binom{6-1}{5-1} (0.7)^5 (1-0.7)^{6-5}$$

$$P(X = 6) = \binom{5}{4} (0.7)^5 (0.3)^1$$

$$P(X = 6) = 5 \cdot (0.16807) \cdot (0.3) = 0.2521$$

There is a 25.21% chance.

Problem 3.5.2 (). (Alternative parameterization). Let Y be the number of *failures* before r successes are achieved. Find the PMF of Y.

Solution: Let X be our original Negative Binomial RV (total trials). $X = Y + r$ (Total trials = failures + successes). Y can take values $y \in \{0, 1, 2, \dots\}$. The corresponding k (total trials) is $k = y + r$. The PMF for X is $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$. We can substitute $k = y + r$:

$$P(Y = y) = P(X = y + r) = \binom{(y+r)-1}{r-1} p^r (1-p)^{(y+r)-r}$$

$$P(Y = y) = \binom{y+r-1}{r-1} p^r (1-p)^y$$

Using the identity $\binom{n}{k} = \binom{n}{n-k}$, we have $\binom{y+r-1}{r-1} = \binom{y+r-1}{(y+r-1)-(r-1)} = \binom{y+r-1}{y}$. So an equivalent PMF is:

$$P(Y = y) = \binom{y+r-1}{y} p^r (1-p)^y, \quad y = 0, 1, 2, \dots$$

This is the standard alternative parameterization for the Negative Binomial distribution, often used in statistics as it resembles the Poisson PMF (and converges to it).

Problem 3.5.3 (). Show that the Negative Binomial distribution can be represented as a Poisson-Gamma mixture. That is, if $Y \sim \text{Poisson}(\lambda)$ and λ itself is a random variable following a $\text{Gamma}(\alpha, \beta)$ distribution, the resulting (marginal) distribution of Y is Negative Binomial.

Solution: This is a hierarchical model. 1. $\lambda \sim \text{Gamma}(\alpha = r, \beta)$ (Using r for shape to match NB) $f(\lambda) = \frac{1}{\Gamma(r)\beta^r} \lambda^{r-1} e^{-\lambda/\beta}$ 2. $(Y|\lambda) \sim \text{Poisson}(\lambda)$ $P(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$

We want to find the marginal PMF $P(Y = y)$ by "integrating out" λ:

$$P(Y = y) = \int_0^\infty P(Y = y|\lambda) f(\lambda)\, d\lambda$$

$$P(Y = y) = \int_0^\infty \left(\frac{e^{-\lambda}\lambda^y}{y!}\right) \left(\frac{1}{\Gamma(r)\beta^r} \lambda^{r-1} e^{-\lambda/\beta}\right) d\lambda$$

Combine terms:

$$P(Y = y) = \frac{1}{y!\Gamma(r)\beta^r} \int_0^\infty \lambda^{y+r-1} e^{-\lambda-\lambda/\beta}\, d\lambda$$

$$P(Y = y) = \frac{1}{y!\Gamma(r)\beta^r} \int_0^\infty \lambda^{(y+r)-1} e^{-\lambda(1+1/\beta)}\, d\lambda$$

$$P(Y = y) = \frac{1}{y!\Gamma(r)\beta^r} \int_0^\infty \lambda^{(y+r)-1} e^{-\lambda\left(\frac{\beta+1}{\beta}\right)}\, d\lambda$$

This integral looks just like the kernel of a Gamma distribution. The PDF of a $\text{Gamma}(A, B)$ is $\frac{1}{\Gamma(A)B^A} x^{A-1} e^{-x/B}$. The integral is $\int_0^\infty x^{A-1} e^{-x/B}\, dx = \Gamma(A)B^A$. In our case, $A = y + r$ and $1/B = \frac{\beta+1}{\beta}$, so $B = \frac{\beta}{\beta+1}$. The integral evaluates to: $\Gamma(y+r)\left(\frac{\beta}{\beta+1}\right)^{y+r}$.

$$P(Y = y) = \frac{1}{y!\Gamma(r)\beta^r} \cdot \Gamma(y+r) \left(\frac{\beta}{\beta+1}\right)^{y+r}$$

$$P(Y = y) = \frac{\Gamma(y+r)}{y!\Gamma(r)} \cdot \frac{1}{\beta^r} \cdot \frac{\beta^{y+r}}{(\beta+1)^{y+r}}$$

$$P(Y = y) = \frac{\Gamma(y+r)}{y!\Gamma(r)} \cdot \frac{\beta^y}{(\beta+1)^{y+r}}$$

$$P(Y = y) = \frac{(y+r-1)!}{y!(r-1)!} \cdot \frac{\beta^y}{(\beta+1)^y(\beta+1)^r}$$

$$P(Y=y) = \binom{y+r-1}{y}\left(\frac{1}{\beta+1}\right)^r\left(\frac{\beta}{\beta+1}\right)^y$$

This is exactly the PMF for the Negative Binomial distribution in the (failures) parameterization $P(Y=y) = \binom{y+r-1}{y}p^r(1-p)^y$. We can match parameters: $p = \frac{1}{\beta+1}$ and $1-p = \frac{\beta}{\beta+1}$. Thus, a Poisson-Gamma mixture creates a Negative Binomial distribution. This is often used in modeling, e.g., if the "rate" λ of an event is not constant but varies according to a Gamma distribution.

3.6 Poisson Distribution

Definition and Properties

A random variable $X \sim \text{Poisson}(\lambda)$ models the number of events occurring in a fixed interval of time or space, given a constant average rate λ.

- **PMF:** $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k \in \{0, 1, 2, \dots\}$
- **PGF:** $G_X(s) = e^{\lambda(s-1)}$
- **MGF:** $M_X(t) = e^{\lambda(e^t-1)}$
- **CF:** $\varphi_X(t) = e^{\lambda(e^{it}-1)}$
- **Mean:** $\mathbb{E}[X] = \lambda$
- **Variance:** $\text{Var}(X) = \lambda$

Discussion: A key property is that the mean equals the variance. It's the limit of the Binomial distribution for rare events.

Solved Problems

Problem 3.6.1 (). A call center receives an average of $\lambda = 3$ calls per minute. What is the probability that they receive exactly 0 calls in a given minute?

Solution: This is a Poisson($\lambda = 3$) process. We want $P(X=0)$. Using the PMF:

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$P(X=0) = \frac{e^{-3}3^0}{0!}$$

Recall that $3^0 = 1$ and $0! = 1$.

$$P(X=0) = e^{-3} \approx 0.0498$$

There is approximately a 4.98% chance of receiving no calls in a minute.

Problem 3.6.2 (). If $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ are independent, find the distribution of $Z = X+Y$.

Solution: We will use MGFs (PGFs or CFs would also work perfectly). The MGF for X is $M_X(t) = e^{\lambda_1(e^t-1)}$. The MGF for Y is $M_Y(t) = e^{\lambda_2(e^t-1)}$. Since X and Y are independent, the MGF of their sum $Z = X+Y$ is the product of their MGFs:

$$M_Z(t) = M_X(t) \cdot M_Y(t)$$

$$M_Z(t) = e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)}$$

Combine the exponents:

$$M_Z(t) = \exp\left(\lambda_1(e^t - 1) + \lambda_2(e^t - 1)\right)$$

$$M_Z(t) = \exp\left((\lambda_1 + \lambda_2)(e^t - 1)\right)$$

This is the MGF of a Poisson distribution with parameter $\lambda = \lambda_1 + \lambda_2$. By the Uniqueness Property of MGFs, Z must follow this distribution. Therefore, $Z \sim \text{Poisson}(\lambda_1 + \lambda_2)$. This is the additive property of the Poisson distribution.

Problem 3.6.3 (). Find the conditional distribution of X given $X + Y = n$, where $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ are independent.

Solution: We want to find $P(X = k|X + Y = n)$ for $k = 0, 1, \ldots, n$. By definition of conditional probability:

$$P(X = k|X + Y = n) = \frac{P(X = k \text{ and } X + Y = n)}{P(X + Y = n)}$$

The event "$(X = k)$ and $(X + Y = n)$" is the same as "$(X = k)$ and $(Y = n - k)$".

$$P(X = k|X + Y = n) = \frac{P(X = k \text{ and } Y = n - k)}{P(X + Y = n)}$$

Since X and Y are independent, $P(X = k, Y = n - k) = P(X = k)P(Y = n - k)$. From the previous problem, we know $Z = X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$. Let's write out the PMFs:

- $P(X = k) = \frac{e^{-\lambda_1}\lambda_1^k}{k!}$
- $P(Y = n - k) = \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}$
- $P(X + Y = n) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}$

Now, substitute these into the fraction:

$$P(X = k|X + Y = n) = \frac{\left(\frac{e^{-\lambda_1}\lambda_1^k}{k!}\right)\left(\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\right)}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}}$$

The $e^{-\lambda_1}$ and $e^{-\lambda_2}$ terms in the numerator combine to $e^{-(\lambda_1+\lambda_2)}$, which cancels with the $e^{-(\lambda_1+\lambda_2)}$ in the denominator.

$$P(X = k|X + Y = n) = \frac{\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}}{\frac{(\lambda_1+\lambda_2)^n}{n!}}$$

Rearrange the terms:

$$P(X = k|X + Y = n) = \frac{n!}{k!(n-k)!} \cdot \frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n}$$

Recognize $\frac{n!}{k!(n-k)!} = \binom{n}{k}$.

$$P(X = k|X + Y = n) = \binom{n}{k}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n}$$

$$P(X = k|X + Y = n) = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}$$

This is the PMF of a Binomial(n, p) distribution, with $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$. So, $(X|X + Y = n) \sim \text{Binomial}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$. This means if we know the total number of events n that occurred from two independent Poisson processes, the distribution of how many came from the first process is Binomial.

3.7 Hypergeometric Distribution

Definition and Properties

The $X \sim \text{Hypergeometric}(N, K, n)$ distribution models sampling *without replacement.* From a population of N items, K are "successes" and $N - K$ are "failures". We draw a sample of n items. X is the number of successes in our sample.

- **PMF:** $P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
- **Support:** $\max(0, n - (N - K)) \le k \le \min(n, K)$
- **PGF/MGF/CF:** No simple closed form.
- **Mean:** $\mathbb{E}[X] = n\frac{K}{N}$
- **Variance:** $\text{Var}(X) = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\left(\frac{N-n}{N-1}\right)$

Discussion: The mean $\mathbb{E}[X] = n(K/N)$ is intuitive: n draws times the probability $p = K/N$ of a success on the first draw. The variance is $np(1-p)$ times the "finite population correction factor" $(N-n)/(N-1)$, which is < 1. This shows that sampling without replacement reduces variance compared to the Binomial (sampling *with* replacement).

Solved Problems

Problem 3.7.1 (). A deck of 52 cards ($N = 52$) has 4 Aces ($K = 4$). If you are dealt a 5-card hand ($n = 5$), what is the probability you get exactly 2 Aces ($k = 2$)?

Solution: This is a Hypergeometric(N=52, K=4, n=5) problem. We want $P(X = 2)$. We need to choose $k = 2$ Aces from the $K = 4$ available Aces: $\binom{K}{k} = \binom{4}{2}$. We must also choose $n - k = 5 - 2 = 3$ non-Aces from the $N - K = 52 - 4 = 48$ available non-Aces: $\binom{N-K}{n-k} = \binom{48}{3}$. The total number of 5-card hands is $\binom{N}{n} = \binom{52}{5}$.

$$P(X = 2) = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}$$

$$P(X = 2) = \frac{\left(\frac{4\cdot 3}{2}\right)\left(\frac{48\cdot 47\cdot 46}{3\cdot 2\cdot 1}\right)}{\left(\frac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5\cdot 4\cdot 3\cdot 2\cdot 1}\right)}$$

$$P(X = 2) = \frac{6 \cdot 17296}{2598960} = \frac{103776}{2598960} \approx 0.0399$$

There is approximately a 3.99% chance.

Problem 3.7.2 (). Show that as $N \to \infty$ while $K/N \to p$ (and n is fixed), the Hypergeometric(N, K, n) distribution converges to the Binomial(n, p) distribution.

Solution: We start with the PMF: $P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$.

$$P(X = k) = \frac{K!}{(K-k)!k!} \cdot \frac{(N-K)!}{(N-K-n+k)!(n-k)!} \cdot \frac{n!(N-n)!}{N!}$$

Let's rewrite the factorial terms as falling products (Pochhammer symbols). $\binom{K}{k} = \frac{K(K-1)\ldots(K-k+1)}{k!}$ $\binom{N-K}{n-k} = \frac{(N-K)\ldots(N-K-n+k+1)}{(n-k)!}$ $\binom{N}{n} = \frac{N(N-1)\ldots(N-n+1)}{n!}$

$$P(X = k) = \frac{n!}{k!(n-k)!} \frac{[K(K-1)\ldots(K-k+1)][(N-K)\ldots(N-K-n+k+1)]}{N(N-1)\ldots(N-n+1)}$$

$$P(X=k)=\binom{n}{k}\frac{(K\cdot(K-1)\dots)\cdot((N-K)\cdot(N-K-1)\dots)}{N\cdot(N-1)\cdot(N-2)\dots}$$

The numerator has k terms from K and $n-k$ terms from $N-K$, for a total of n terms. The denominator also has n terms. Now, let $N\to\infty$ and $K\to\infty$ such that $K/N\to p$. This implies $(N-K)/N\to 1-p$. Let's divide every term in the fraction by N:

$$\frac{K}{N}\to p$$

$$\frac{K-1}{N-1}=\frac{K/N-1/N}{1-1/N}\to\frac{p-0}{1-0}=p$$

$$\frac{N-K}{N-k}=\frac{(N-K)/N}{1-k/N}\to\frac{1-p}{1-0}=1-p$$

$$\frac{N-K-1}{N-k-1}=\frac{(N-K)/N-1/N}{1-(k+1)/N}\to\frac{1-p}{1}=1-p$$

In general, for any fixed j, $\frac{K-j}{N-j}\to p$ and $\frac{N-K-j}{N-k-j}\to 1-p$. Let's group the n terms in the numerator and n terms in the denominator:

$$P(X=k)=\binom{n}{k}\left[\frac{K}{N}\cdot\frac{K-1}{N-1}\dots\right]\left[\frac{N-K}{N-k}\dots\right]$$

$$P(X=k)=\binom{n}{k}\underbrace{\left(\frac{K}{N}\frac{K-1}{N-1}\dots\frac{K-k+1}{N-k+1}\right)}_{k\text{ terms}}\underbrace{\left(\frac{N-K}{N-k}\frac{N-K-1}{N-k-1}\dots\frac{N-K-n+k+1}{N-n+1}\right)}_{n-k\text{ terms}}$$

As $N\to\infty$, every one of the k terms in the first group converges to p. As $N\to\infty$, every one of the $n-k$ terms in the second group converges to $(1-p)$.

$$\lim_{N\to\infty}P(X=k)=\binom{n}{k}(p\cdot p\dots p)\cdot((1-p)\cdot(1-p)\dots(1-p))$$

$$\lim_{N\to\infty}P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

This is the PMF of the Binomial(n,p) distribution. This makes intuitive sense: when the population N is very large, drawing n items *without* replacement is "almost" the same as drawing *with* replacement, because the probability $p=K/N$ barely changes after each draw.

Problem 3.7.3 (). Derive the mean $\mathbb{E}[X]$ of the Hypergeometric(N,K,n) distribution using the "indicator variable" method.

Solution: Let X be the total number of successes in the n draws. We can write X as the sum of n indicator variables. Let I_i be an indicator variable for the i-th draw being a success, for $i=1,\dots,n$. $I_i=1$ if the i-th draw is a success, $I_i=0$ otherwise. Then $X=\sum_{i=1}^n I_i$. By the linearity of expectation, $\mathbb{E}[X]=\mathbb{E}\left[\sum_{i=1}^n I_i\right]=\sum_{i=1}^n\mathbb{E}[I_i]$. The expectation of an indicator variable is the probability of the event it indicates: $\mathbb{E}[I_i]=P(I_i=1)$. $P(I_i=1)$ is the probability that the i-th draw is a success.

- For $i=1$: $P(I_1=1)=K/N$. This is the probability of drawing a success on the first try. $\mathbb{E}[I_1]=K/N$.
- For $i=2$: $P(I_2=1)$. We can find this using the law of total probability, but a simpler argument is based on symmetry. Any draw is equally likely to be any of the N items. K of these items are successes. Therefore, the probability that a "random" draw (like the i-th draw) is a success must be K/N.
- Formal check for $i=2$: $P(I_2=1)=P(I_2=1|I_1=1)P(I_1=1)+P(I_2=1|I_1=0)P(I_1=0)$ $P(I_2=1)=\left(\frac{K-1}{N-1}\right)\left(\frac{K}{N}\right)+\left(\frac{K}{N-1}\right)\left(\frac{N-K}{N}\right)$ $P(I_2=1)=\frac{K(K-1)+K(N-K)}{N(N-1)}=\frac{K^2-K+KN-K^2}{N(N-1)}=\frac{KN-K}{N(N-1)}=\frac{K(N-1)}{N(N-1)}=\frac{K}{N}$

This symmetry argument holds for all i. The marginal probability of any single draw i being a success is K/N.

$$P(I_i=1)=K/N\quad\text{for all } i=1,\dots,n$$

Therefore, $\mathbb{E}[I_i]=K/N$ for all i. Now we compute the sum:

$$\mathbb{E}[X]=\sum_{i=1}^n\mathbb{E}[I_i]=\sum_{i=1}^n\frac{K}{N}=n\frac{K}{N}$$

This is a much more elegant derivation of the mean than using the PMF and combinatorics.

3.8 Multivariate Hypergeometric Distribution

Definition and Properties

This generalizes the Hypergeometric to r categories. A population of N items has K_i items of type i, for $i = 1, \dots, r$, with $\sum K_i = N$. We draw a sample of n items without replacement. The RV $\mathbf{X} = (X_1, \dots, X_r)$ counts the number of items of each type in the sample.

- **PMF:** $P(X_1 = k_1, \dots, X_r = k_r) = \frac{\binom{K_1}{k_1}\binom{K_2}{k_2}\dots\binom{K_r}{k_r}}{\binom{N}{n}}$, where $\sum k_i = n$.
- **Mean:** $\mathbb{E}[X_i] = n\frac{K_i}{N}$
- **Variance:** $\text{Var}(X_i) = n\frac{K_i}{N}(1 - \frac{K_i}{N})(\frac{N-n}{N-1})$
- **Covariance:** $\text{Cov}(X_i, X_j) = -n\frac{K_i K_j}{N^2}(\frac{N-n}{N-1})$ for $i \neq j$.

Discussion: The marginal distribution of any single component X_i is Hypergeometric(N, K_i, n). The properties are direct analogues of the Multinomial, but with the finite population correction.

Solved Problems

Problem 3.8.1 (). A bag contains 20 marbles ($N = 20$): 10 Red ($K_1 = 10$), 7 Blue ($K_2 = 7$), 3 Green ($K_3 = 3$). You draw 5 marbles ($n = 5$). What is the probability you get 2 Red, 2 Blue, and 1 Green?

Solution: We want $P(X_1 = 2, X_2 = 2, X_3 = 1)$. Check: $k_1 + k_2 + k_3 = 2 + 2 + 1 = 5 = n$. We use the PMF:

$$P(\mathbf{X} = (2, 2, 1)) = \frac{\binom{K_1}{k_1}\binom{K_2}{k_2}\binom{K_3}{k_3}}{\binom{N}{n}}$$

$$P = \frac{\binom{10}{2}\binom{7}{2}\binom{3}{1}}{\binom{20}{5}}$$

$$P = \frac{\left(\frac{10\cdot 9}{2}\right)\left(\frac{7\cdot 6}{2}\right)(3)}{\left(\frac{20\cdot 19\cdot 18\cdot 17\cdot 16}{5\cdot 4\cdot 3\cdot 2\cdot 1}\right)}$$

$$P = \frac{45 \cdot 21 \cdot 3}{15504} = \frac{2835}{15504} \approx 0.1829$$

There is approximately an 18.29% chance.

Problem 3.8.2 (). A committee of $n = 10$ people is to be chosen from a group of $N = 40$ people: 15 men ($K_1 = 15$) and 25 women ($K_2 = 25$). Let X_1 be the number of men and X_2 be the number of women on the committee. Find $\mathbb{E}[X_1]$ and $\text{Var}(X_1)$.

Solution: This is a Multivariate Hypergeometric distribution with $r = 2$. This is equivalent to the standard (univariate) Hypergeometric distribution. $X_1 \sim$ Hypergeometric$(N = 40, K = K_1 = 15, n = 10)$. The mean is:

$$\mathbb{E}[X_1] = n\frac{K_1}{N} = 10 \cdot \frac{15}{40} = \frac{150}{40} = 3.75$$

We expect 3.75 men on the committee. The variance is:

$$\text{Var}(X_1) = n\frac{K_1}{N}\left(1 - \frac{K_1}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Let's plug in the values. $p = K_1/N = 15/40 = 3/8$ $1 - p = 25/40 = 5/8$ $n = 10$ Finite Population Correction (FPC): $\frac{N-n}{N-1} = \frac{40-10}{40-1} = \frac{30}{39} = \frac{10}{13}$

$$\text{Var}(X_1) = 10 \cdot \left(\frac{3}{8}\right) \cdot \left(\frac{5}{8}\right) \cdot \left(\frac{10}{13}\right)$$

$$\text{Var}(X_1) = \frac{10 \cdot 15 \cdot 10}{8 \cdot 8 \cdot 13} = \frac{1500}{832} \approx 1.803$$

Problem 3.8.3 (). Derive the covariance $\text{Cov}(X_i, X_j)$ for $i \neq j$ in the Multivariate Hypergeometric distribution, using indicator variables.

Solution: Let $X_i = \sum_{k=1}^{n} I_{i,k}$ where $I_{i,k} = 1$ if draw k is of type i, 0 otherwise. Let $X_j = \sum_{l=1}^{n} I_{j,l}$ where $I_{j,l} = 1$ if draw l is of type j, 0 otherwise.

$$\text{Cov}(X_i, X_j) = \text{Cov}\left(\sum_{k=1}^{n} I_{i,k}, \sum_{l=1}^{n} I_{j,l}\right) = \sum_{k=1}^{n}\sum_{l=1}^{n} \text{Cov}(I_{i,k}, I_{j,l})$$

We split this sum into two cases: $k = l$ and $k \neq l$. **Case 1:** $k = l$ We need $\text{Cov}(I_{i,k}, I_{j,k}) = \mathbb{E}[I_{i,k}I_{j,k}] - \mathbb{E}[I_{i,k}]\mathbb{E}[I_{j,k}]$. The product $I_{i,k}I_{j,k}$ is 1 if and only if draw k is *both* type i and type j. This is impossible, as the categories are distinct. So $I_{i,k}I_{j,k} = 0$, which means $\mathbb{E}[I_{i,k}I_{j,k}] = 0$. By symmetry, $\mathbb{E}[I_{i,k}] = p_i = K_i/N$ and $\mathbb{E}[I_{j,k}] = p_j = K_j/N$. $\text{Cov}(I_{i,k}, I_{j,k}) = 0 - (K_i/N)(K_j/N) = -K_iK_j/N^2$. There are n such terms (for $k = l = 1, \ldots, n$). **Case 2:** $k \neq l$ We need $\text{Cov}(I_{i,k}, I_{j,l}) = \mathbb{E}[I_{i,k}I_{j,l}] - \mathbb{E}[I_{i,k}]\mathbb{E}[I_{j,l}]$. $\mathbb{E}[I_{i,k}I_{j,l}] = P(I_{i,k} = 1 \text{ and } I_{j,l} = 1) = P(I_{j,l} = 1 | I_{i,k} = 1)P(I_{i,k} = 1)$. $P(I_{i,k} = 1) = K_i/N$. $P(I_{j,l} = 1 | I_{i,k} = 1)$ is the probability draw l is type j, given draw k was type i. Since $k \neq l$, this is a draw from the remaining $N - 1$ items, of which K_j are still type j. So $P = K_j/(N-1)$. $\mathbb{E}[I_{i,k}I_{j,l}] = \frac{K_j}{N-1} \cdot \frac{K_i}{N} = \frac{K_iK_j}{N(N-1)}$. $\mathbb{E}[I_{i,k}]\mathbb{E}[I_{j,l}] = (K_i/N)(K_j/N) = K_iK_j/N^2$. $\text{Cov}(I_{i,k}, I_{j,l}) = \frac{K_iK_j}{N(N-1)} - \frac{K_iK_j}{N^2} = \frac{K_iK_j}{N}\left(\frac{1}{N-1} - \frac{1}{N}\right) = \frac{K_iK_j}{N}\left(\frac{N-(N-1)}{N(N-1)}\right) = \frac{K_iK_j}{N}\left(\frac{1}{N(N-1)}\right) = \frac{K_iK_j}{N^2(N-1)}$. There are $n(n-1)$ such terms (all pairs $k \neq l$). **Combining:**

$$\text{Cov}(X_i, X_j) = \sum_{k=l} \text{Cov}(I_{i,k}, I_{j,k}) + \sum_{k \neq l} \text{Cov}(I_{i,k}, I_{j,l})$$

$$\text{Cov}(X_i, X_j) = n\left(\frac{-K_iK_j}{N^2}\right) + n(n-1)\left(\frac{K_iK_j}{N^2(N-1)}\right)$$

$$\text{Cov}(X_i, X_j) = \frac{-nK_iK_j}{N^2} + \frac{n(n-1)K_iK_j}{N^2(N-1)}$$

$$\text{Cov}(X_i, X_j) = \frac{nK_iK_j}{N^2}\left(-1 + \frac{n-1}{N-1}\right)$$

$$\text{Cov}(X_i, X_j) = \frac{nK_iK_j}{N^2}\left(\frac{-(N-1) + (n-1)}{N-1}\right)$$

$$\text{Cov}(X_i, X_j) = \frac{nK_iK_j}{N^2}\left(\frac{-N+1+n-1}{N-1}\right)$$

$$\text{Cov}(X_i, X_j) = \frac{nK_iK_j}{N^2}\left(\frac{n-N}{N-1}\right) = -n\frac{K_iK_j}{N^2}\left(\frac{N-n}{N-1}\right)$$

This matches the formula. The negative sign confirms that as X_i increases, X_j tends to decrease.

---

4 Continuous Distributions

We now present standard continuous distributions.

4.1 Uniform Distribution

Definition and Properties

A random variable $X \sim \text{Uniform}(a, b)$ has equal probability density for all values in the interval $[a, b]$.

- **PDF:** $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, and 0 otherwise.
- **MGF:** $M_X(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$ for $t \neq 0$, and 1 for $t = 0$.
- **CF:** $\varphi_X(t) = \frac{e^{itb}-e^{ita}}{it(b-a)}$
- **Mean:** $\mathbb{E}[X] = \frac{a+b}{2}$
- **Variance:** $\text{Var}(X) = \frac{(b-a)^2}{12}$

Discussion: This is the simplest continuous distribution, modeling a "completely random" choice from an interval.

Solved Problems

Problem 4.1.1 (). A bus arrives at a stop every 15 minutes. If you arrive at the stop at a random time, what is the probability you have to wait more than 5 minutes?

Solution: Your waiting time X is a random variable. The "random time" of arrival means your wait time is uniformly distributed on the interval $[0, 15]$. $X \sim \text{Uniform}(a = 0, b = 15)$. The PDF is $f(x) = 1/(15-0) = 1/15$ for $0 \leq x \leq 15$. We want to find $P(X > 5)$.

$$P(X > 5) = \int_5^{15} f(x)\,dx = \int_5^{15} \frac{1}{15}\,dx$$

$$P(X > 5) = \frac{1}{15}[x]_5^{15} = \frac{1}{15}(15-5) = \frac{10}{15} = \frac{2}{3}$$

There is a 2/3 chance you wait more than 5 minutes.

Problem 4.1.2 (). Let $X \sim \text{Uniform}(0, 1)$. Find the PDF of the random variable $Y = -\log(X)$.

Solution: This is a transformation of variables. We can use the CDF method. $Y = g(X) = -\log(X)$. Since $X \in (0, 1)$, $Y \in (0, \infty)$. Let $F_Y(y)$ be the CDF of Y. We want $P(Y \leq y)$ for $y > 0$.

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(-\log(X) \leq y) \\ &= P(\log(X) \geq -y) \\ &= P(X \geq e^{-y}) \end{aligned}$$

Since $X \sim \text{Uniform}(0, 1)$, its CDF is $F_X(x) = x$ for $x \in [0, 1]$. $P(X \geq e^{-y}) = 1 - P(X < e^{-y}) = 1 - F_X(e^{-y})$. Since $y > 0$, $e^{-y} \in (0, 1)$, so $F_X(e^{-y}) = e^{-y}$.

$$F_Y(y) = 1 - e^{-y} \quad \text{for } y > 0$$

This is the CDF of Y. To find the PDF $f_Y(y)$, we differentiate the CDF:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}(1 - e^{-y}) = -(-e^{-y}) = e^{-y}$$

This PDF $f_Y(y) = 1 \cdot e^{-1 \cdot y}$ is the PDF of an Exponential($\lambda = 1$) distribution. Therefore, $Y \sim \text{Exponential}(1)$. This is a fundamental result used in simulation, known as the "inverse transform sampling" method for the exponential distribution.

Problem 4.1.3 (). Let $X \sim \text{Uniform}(0,1)$ and $Y \sim \text{Uniform}(0,1)$ be independent. Find the PDF of $Z = X + Y$.

Solution: X and Y have PDF $f(t) = 1$ for $t \in [0,1]$ and 0 otherwise. The PDF of $Z = X + Y$ is the convolution of their individual PDFs:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx$$

$f_X(x) = 1$ for $0 \le x \le 1$. $f_Y(z-x) = 1$ for $0 \le z-x \le 1$, which means $z-1 \le x \le z$. The integrand $f_X(x)f_Y(z-x)$ is 1 only when *both* conditions are met: $0 \le x \le 1$ $z-1 \le x \le z$ The range of Z is $[0,2]$. We must split the problem into cases based on z. **Case 1:** $0 \le z < 1$ The integration bounds on x are $0 \le x \le 1$ and $z-1 \le x \le z$. Since $z < 1$, $z-1 < 0$. The common interval is $0 \le x \le z$.

$$f_Z(z) = \int_0^z 1 \cdot 1\,dx = [x]_0^z = z$$

Case 2: $1 \le z \le 2$ The integration bounds on x are $0 \le x \le 1$ and $z-1 \le x \le z$. Since $z \ge 1$, $z-1 \ge 0$. Since $z \le 2$, $z-1 \le 1$. The common interval is $z-1 \le x \le 1$.

$$f_Z(z) = \int_{z-1}^1 1 \cdot 1\,dx = [x]_{z-1}^1 = 1-(z-1) = 2-z$$

Case 3: $z < 0$ **or** $z > 2$ The intervals do not overlap. The integral is 0. Combining these, we get the PDF for Z:

$$f_Z(z) = \begin{cases} z & \text{for } 0 \le z < 1 \\ 2-z & \text{for } 1 \le z \le 2 \\ 0 & \text{otherwise} \end{cases}$$

This is the "Triangular distribution" (specifically, the Bates distribution for $n = 2$).

4.2 Exponential Distribution

Definition and Properties

A random variable $X \sim \text{Exponential}(\lambda)$ models the "waiting time" between events in a Poisson process.

- **PDF:** $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$.
- **MGF:** $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.
- **CF:** $\varphi_X(t) = \frac{\lambda}{\lambda - it}$
- **Mean:** $\mathbb{E}[X] = 1/\lambda$
- **Variance:** $\text{Var}(X) = 1/\lambda^2$

Discussion: This is the continuous analogue of the Geometric distribution. It possesses the continuous **memoryless property**: $P(X > s+t|X > s) = P(X > t)$.

Solved Problems

Problem 4.2.1 (). The lifetime of a lightbulb follows an exponential distribution with a mean of 1000 hours. What is the probability that a bulb lasts more than 1000 hours?

Solution: The mean is $\mathbb{E}[X] = 1/\lambda = 1000$. This means $\lambda = 1/1000$. The PDF is $f(x) = (1/1000)e^{-x/1000}$. We want $P(X > 1000)$.

$$P(X > 1000) = \int_{1000}^{\infty} \frac{1}{1000} e^{-x/1000}\,dx$$

Let $u = x/1000$, $du = dx/1000$. When $x = 1000, u = 1$.

$$P(X > 1000) = \int_1^\infty e^{-u}\,du = [-e^{-u}]_1^\infty = (0) - (-e^{-1}) = e^{-1}$$

$$P(X > 1000) \approx 0.3679$$

There is a 36.79% chance. This is true for *any* exponential distribution: the probability of lasting longer than the mean is always e^{-1}.

Problem 4.2.2 (). Prove the memoryless property of the Exponential distribution: $P(X > s+t|X > s) = P(X > t)$ for $s, t > 0$.

Solution: First, we find the survival function $P(X > k)$.

$$P(X > k) = \int_k^\infty \lambda e^{-\lambda x}\,dx = \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_k^\infty = [-e^{-\lambda x}]_k^\infty$$

$$P(X > k) = (\lim_{x\to\infty} -e^{-\lambda x}) - (-e^{-\lambda k}) = 0 + e^{-\lambda k} = e^{-\lambda k}$$

Now, we analyze the conditional probability:

$$P(X > s+t|X > s) = \frac{P(X > s+t \text{ and } X > s)}{P(X > s)}$$

As in the geometric case, the event $(X > s+t)$ is a subset of $(X > s)$. Their intersection is just $(X > s+t)$.

$$P(X > s+t|X > s) = \frac{P(X > s+t)}{P(X > s)}$$

Substitute the survival function:

$$P(X > s+t|X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{e^{-\lambda s}e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda t}$$

We also know that $P(X > t) = e^{-\lambda t}$. Therefore, $P(X > s+t|X > s) = P(X > t)$. This means that if the lightbulb has already lasted $s = 1000$ hours, the probability that it lasts *another* $t = 500$ hours is the same as the probability that a brand new bulb lasts 500 hours. The bulb "forgets" its age.

Problem 4.2.3 (). Let $X \sim \text{Exponential}(\lambda_1)$ and $Y \sim \text{Exponential}(\lambda_2)$ be independent. Find the distribution of $Z = \min(X, Y)$.

Solution: We can find the distribution of Z by first finding its CDF, $F_Z(z) = P(Z \le z)$. It's easier to work with the survival function $P(Z > z)$.

$$P(Z > z) = P(\min(X, Y) > z)$$

The minimum of two numbers is greater than z if and only if *both* numbers are greater than z.

$$P(Z > z) = P(X > z \text{ and } Y > z)$$

Since X and Y are independent:

$$P(Z > z) = P(X > z)P(Y > z)$$

From the previous problem, we know the survival function for an exponential is $P(X > z) = e^{-\lambda_1 z}$ and $P(Y > z) = e^{-\lambda_2 z}$.

$$P(Z > z) = (e^{-\lambda_1 z})(e^{-\lambda_2 z}) = e^{-(\lambda_1+\lambda_2)z}$$

This is the survival function of Z. Let's find the CDF:

$$F_Z(z) = P(Z \le z) = 1 - P(Z > z) = 1 - e^{-(\lambda_1+\lambda_2)z}$$

This is the CDF of an exponential distribution. To find the PDF, we differentiate:

$$f_Z(z) = \frac{d}{dz}(1 - e^{-(\lambda_1+\lambda_2)z}) = -(-(\lambda_1 + \lambda_2))e^{-(\lambda_1+\lambda_2)z}$$

$$f_Z(z) = (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)z}$$

This is the PDF of an Exponential distribution with rate $\lambda = \lambda_1 + \lambda_2$. Therefore, $Z = \min(X, Y) \sim \text{Exponential}(\lambda_1 + \lambda_2)$. This is a key result in reliability and stochastic processes: the time until the *first* failure in a system with two independent components is exponentially distributed with a rate equal to the sum of the individual failure rates.

4.3 Gamma Distribution

Definition and Properties

The $X \sim \text{Gamma}(\alpha, \beta)$ (shape $\alpha > 0$, scale $\beta > 0$) distribution is a flexible two-parameter family.

- **PDF:** $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$ for $x > 0$. ($\Gamma(\alpha)$ is the Gamma function).
- **MGF:** $M_X(t) = (1 - \beta t)^{-\alpha}$ for $t < 1/\beta$.
- **CF:** $\varphi_X(t) = (1 - i\beta t)^{-\alpha}$
- **Mean:** $\mathbb{E}[X] = \alpha\beta$
- **Variance:** $\text{Var}(X) = \alpha\beta^2$

Discussion:

- If $\alpha = 1$, Gamma(1, β) $\sim$ Exponential($\lambda = 1/\beta$).
- If $\alpha = n$ (an integer), it is the **Erlang(n, 1/β)** distribution, modeling the sum of n i.i.d. Exponential($\lambda = 1/\beta$) RVs.
- If $\alpha = \nu/2$ and $\beta = 2$, it is the Chi-square(ν) distribution.

Solved Problems

Problem 4.3.1 (). Given $X \sim \text{Gamma}(\alpha = 3, \beta = 2)$. Find its mean, variance, and MGF.

Solution: This is a direct application of the formulas.

- **Mean:** $\mathbb{E}[X] = \alpha\beta = 3 \cdot 2 = 6$.
- **Variance:** $\text{Var}(X) = \alpha\beta^2 = 3 \cdot (2^2) = 12$.
- **MGF:** $M_X(t) = (1 - \beta t)^{-\alpha} = (1 - 2t)^{-3}$, valid for $t < 1/2$.

Problem 4.3.2 (). Use MGFs to show that if $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \sim \text{Gamma}(\alpha_2, \beta)$ are independent, then $Z = X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$. (Additive property for same scale).

Solution: The MGF for X is $M_X(t) = (1 - \beta t)^{-\alpha_1}$. The MGF for Y is $M_Y(t) = (1 - \beta t)^{-\alpha_2}$. Note: This property only holds if the scale parameter β is the same. Since X and Y are independent, the MGF of their sum $Z = X + Y$ is the product of their MGFs:

$$M_Z(t) = M_X(t) \cdot M_Y(t)$$

$$M_Z(t) = (1 - \beta t)^{-\alpha_1} \cdot (1 - \beta t)^{-\alpha_2}$$

Combine the exponents:

$$M_Z(t) = (1 - \beta t)^{-(\alpha_1+\alpha_2)}$$

This is the MGF of a Gamma distribution with shape $\alpha = \alpha_1 + \alpha_2$ and scale β. By the Uniqueness Property of MGFs, Z must follow this distribution. Therefore, $Z \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$. This confirms that the sum of n i.i.d. Exponential(λ) $\sim$ Gamma$(1, 1/\lambda)$ variables is Gamma$(n, 1/\lambda)$.

Problem 4.3.3 (). Let $X \sim \text{Gamma}(\alpha, \beta)$. Find the MGF of $Y = cX$ for some constant $c > 0$.

Solution: We want to find $M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(cX)}] = \mathbb{E}[e^{(ct)X}]$. This is, by definition, the MGF of X, $M_X(t)$, evaluated at the point (ct). The MGF of X is $M_X(t) = (1 - \beta t)^{-\alpha}$. Therefore:

$$M_Y(t) = M_X(ct) = (1 - \beta(ct))^{-\alpha} = (1 - (\beta c)t)^{-\alpha}$$

We recognize this as the MGF of a Gamma distribution. It has shape parameter $\alpha' = \alpha$ and scale parameter $\beta' = \beta c$. So, $Y = cX \sim \text{Gamma}(\alpha, c\beta)$. This is the "scaling property" of the Gamma distribution. Multiplying a Gamma RV by a constant c scales the scale parameter β by c, which makes intuitive sense. The mean also scales: $\mathbb{E}[Y] = \alpha\beta' = \alpha(c\beta) = c(\alpha\beta) = c\mathbb{E}[X]$.

4.4 Normal (Gaussian) Distribution

Definition and Properties

The $X \sim N(\mu, \sigma^2)$ distribution is the most common distribution in statistics, characterized by its "bell curve."

- **PDF:** $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
- **MGF:** $M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$
- **CF:** $\varphi_X(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$
- **Mean:** $\mathbb{E}[X] = \mu$
- **Variance:** $\text{Var}(X) = \sigma^2$

Discussion: It is the limiting distribution of sums of i.i.d. variables (CLT). Any linear combination of independent Normal RVs is also a Normal RV.

Solved Problems

Problem 4.4.1 (). Assume IQ scores are normally distributed with a mean $\mu = 100$ and standard deviation $\sigma = 15$. (So $X \sim N(100, 15^2)$). What is the probability a randomly selected person has an IQ between 85 and 115?

Solution: We want $P(85 \le X \le 115)$. We "standardize" these values to Z-scores, where $Z = (X - \mu)/\sigma \sim N(0, 1)$. Lower bound: $Z_1 = (85 - 100)/15 = -15/15 = -1$. Upper bound: $Z_2 = (115 - 100)/15 = 15/15 = +1$.

$$P(85 \le X \le 115) = P(-1 \le Z \le 1)$$

By symmetry, $P(-1 \le Z \le 1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1$. From a standard normal Z-table, $\Phi(1) \approx 0.8413$.

$$P(-1 \le Z \le 1) = 0.8413 - (1 - 0.8413) = 0.8413 - 0.1587 = 0.6826$$

This is the "68-95-99.7" rule: approximately 68.3% of the data lies within one standard deviation of the mean.

Problem 4.4.2 (). Let $X \sim N(\mu, \sigma^2)$. Find the MGF of $Y = aX + b$.

Solution: We want to find $M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(aX+b)}] = \mathbb{E}[e^{(at)X+bt}]$.

$$M_Y(t) = e^{bt}\mathbb{E}[e^{(at)X}]$$

The term $\mathbb{E}[e^{(at)X}]$ is the MGF of X, $M_X(t)$, evaluated at the point (at). The MGF of X is $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$. So, $M_X(at) = \exp(\mu(at) + \frac{1}{2}\sigma^2(at)^2) = \exp(a\mu t + \frac{1}{2}a^2\sigma^2t^2)$. Substitute this back:

$$M_Y(t) = e^{bt} \cdot \exp\left(a\mu t + \frac{1}{2}a^2\sigma^2 t^2\right)$$

$$M_Y(t) = \exp\left(bt + a\mu t + \frac{1}{2}(a^2\sigma^2)t^2\right)$$

$$M_Y(t) = \exp\left((a\mu + b)t + \frac{1}{2}(a\sigma)^2 t^2\right)$$

This is the MGF of a Normal distribution with:

- Mean $\mu_Y = a\mu + b$
- Variance $\sigma_Y^2 = (a\sigma)^2 = a^2\sigma^2$

By Uniqueness, $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$. This shows that any linear transformation of a Normal RV is also a Normal RV.

Problem 4.4.3 (). Prove the Box-Muller Transform. Let $U_1, U_2 \sim \text{Uniform}(0,1)$ be independent. Show that $Z_1 = \sqrt{-2\log U_1}\cos(2\pi U_2)$ and $Z_2 = \sqrt{-2\log U_1}\sin(2\pi U_2)$ are independent $N(0,1)$ random variables.

Solution: This is a multivariate change of variables. We have the inverse transformation: $R^2 = Z_1^2 + Z_2^2 = (-2\log U_1)(\cos^2(2\pi U_2) + \sin^2(2\pi U_2)) = -2\log U_1$. So $U_1 = \exp(-(Z_1^2 + Z_2^2)/2)$. And $\frac{Z_2}{Z_1} = \tan(2\pi U_2)$, so $\Theta = 2\pi U_2 = \arctan(Z_2/Z_1)$. Thus $U_2 = \frac{1}{2\pi}\arctan(Z_2/Z_1)$. We need the Jacobian of this transformation $(U_1, U_2) \to (Z_1, Z_2)$. It is simpler to find the Jacobian of $(Z_1, Z_2) \to (U_1, U_2)$ and take its reciprocal. Let's use polar coordinates $R = \sqrt{-2\log U_1}$ and $\Theta = 2\pi U_2$. $Z_1 = R\cos\Theta$ $Z_2 = R\sin\Theta$ $U_1 = e^{-R^2/2}$ $U_2 = \Theta/(2\pi)$ The joint PDF of (U_1, U_2) is $f(u_1, u_2) = 1$ for $u_1, u_2 \in (0,1)$. The joint PDF of (R, Θ) is found via the Jacobian: $R \in (0, \infty)$, $\Theta \in (0, 2\pi)$. $J = \det\begin{vmatrix} \partial u_1/\partial r & \partial u_1/\partial\theta \\ \partial u_2/\partial r & \partial u_2/\partial\theta \end{vmatrix} = \det\begin{vmatrix} -re^{-r^2/2} & 0 \\ 0 & 1/(2\pi) \end{vmatrix} = -\frac{r}{2\pi}e^{-r^2/2}$ The PDF of (R, Θ) is $f(r, \theta) = f(u_1(r,\theta), u_2(r,\theta))|J| = 1 \cdot \frac{r}{2\pi}e^{-r^2/2}$. Now we transform from (R, Θ) to (Z_1, Z_2). $f(z_1, z_2) = f(r(z_1, z_2), \theta(z_1, z_2))|J_{(R,\Theta)\to(Z_1,Z_2)}|^{-1}$ The Jacobian of polar transformation is $J = r$. So $|J|^{-1} = 1/r$. $f(z_1, z_2) = f(r, \theta) \cdot (1/r) = \left(\frac{r}{2\pi}e^{-r^2/2}\right)\frac{1}{r} = \frac{1}{2\pi}e^{-r^2/2}$ Substitute $r^2 = z_1^2 + z_2^2$:

$$f(z_1, z_2) = \frac{1}{2\pi}e^{-(z_1^2+z_2^2)/2}$$

We can factor this joint PDF:

$$f(z_1, z_2) = \left(\frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}\right) \cdot \left(\frac{1}{\sqrt{2\pi}}e^{-z_2^2/2}\right)$$

$$f(z_1, z_2) = f_{Z_1}(z_1) \cdot f_{Z_2}(z_2)$$

This shows that the joint PDF is the product of two $N(0,1)$ PDFs. Therefore, Z_1 and Z_2 are independent $N(0,1)$ random variables.

4.5 Lognormal Distribution

Definition and Properties

A random variable X is Lognormal(μ, σ^2) if $Y = \log(X)$ is Normal(μ, σ^2).

- **PDF:** $f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$ for $x > 0$.
- **MGF:** Does not exist (all moments are finite, but it grows too fast).
- **CF:** No simple closed form.
- **Mean:** $\mathbb{E}[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$
- **Variance:** $\mathrm{Var}(X) = \left(e^{\sigma^2} - 1\right) \exp\left(2\mu + \sigma^2\right)$

Discussion: Used to model quantities that are products of many small factors (like the Normal models sums). Common in finance and biology.

Solved Problems

Problem 4.5.1 (). If $X \sim$ Lognormal$(\mu = 1, \sigma^2 = 4)$, find $\mathbb{E}[X]$.

Solution: This is a direct application of the formula.

$$\mathbb{E}[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right) = \exp\left(1 + \frac{1}{2}(4)\right) = \exp(1+2) = e^3$$

$\mathbb{E}[X] \approx 20.086$. Note that the mean is *not* $e^\mu = e^1$.

Problem 4.5.2 (). Let $Y \sim N(\mu, \sigma^2)$. Let $X = e^Y$. Find the k-th raw moment of X, $\mathbb{E}[X^k]$. Use this to derive the mean and variance.

Solution: $X \sim$ Lognormal(μ, σ^2). We want $\mathbb{E}[X^k] = \mathbb{E}[(e^Y)^k] = \mathbb{E}[e^{kY}]$. This is, by definition, the MGF of Y, $M_Y(t)$, evaluated at $t = k$. $Y \sim N(\mu, \sigma^2)$, so its MGF is $M_Y(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$. Therefore:

$$\mathbb{E}[X^k] = M_Y(k) = \exp\left(\mu k + \frac{1}{2}\sigma^2 k^2\right)$$

This is the formula for the k-th moment. **Derive Mean ($\mathbb{E}[X]$):** Set $k = 1$:

$$\mathbb{E}[X] = \exp\left(\mu(1) + \frac{1}{2}\sigma^2(1)^2\right) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

This matches the formula. **Derive Variance (Var(X)):** $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. We need $\mathbb{E}[X^2]$. Set $k = 2$:

$$\mathbb{E}[X^2] = \exp\left(\mu(2) + \frac{1}{2}\sigma^2(2)^2\right) = \exp\left(2\mu + 2\sigma^2\right)$$

Now, plug in:

$$\mathrm{Var}(X) = \exp(2\mu + 2\sigma^2) - \left[\exp(\mu + \frac{1}{2}\sigma^2)\right]^2$$

$$\mathrm{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$$

Factor out the common term $\exp(2\mu + \sigma^2)$:

$$\mathrm{Var}(X) = \exp(2\mu + \sigma^2)\left[\exp(\sigma^2) - 1\right]$$

This matches the formula.

Problem 4.5.3 (). Let $X_1, X_2, \ldots, X_n$ be i.i.d. Lognormal(μ, σ^2) variables. Find the distribution of the product $P = \prod_{i=1}^n X_i$.

Solution: P is a Lognormal variable. We can show this by looking at $\log(P)$.

$$\log(P) = \log\left(\prod_{i=1}^n X_i\right) = \sum_{i=1}^n \log(X_i)$$

Let $Y_i = \log(X_i)$. By definition of the Lognormal distribution, $Y_i \sim N(\mu, \sigma^2)$. We have a sum of n i.i.d. Normal random variables:

$$\log(P) = \sum_{i=1}^n Y_i$$

Let $W = \log(P)$. From the properties of Normal distributions, the sum W is also normally distributed. The mean of W is $\mathbb{E}[W] = \sum \mathbb{E}[Y_i] = \sum \mu = n\mu$. The variance of W (by independence) is $\mathrm{Var}(W) = \sum \mathrm{Var}(Y_i) = \sum \sigma^2 = n\sigma^2$. So, $W = \log(P) \sim N(n\mu, n\sigma^2)$. Since the logarithm of P is normally distributed, P is, by definition, lognormally distributed. $P \sim \text{Lognormal}(\mu_P = n\mu, \sigma_P^2 = n\sigma^2)$.
