1 Moments

The moments and central moments are numbers that encode the general shape and behavior of a random variable.

Definition 1 (Moments)**.** For each positive integer n, the nth *moment* of X is the number $\mathbb{E}[X^n]$.

The most common moment is the first one, which we simply call the expected value. This measures the typical value of the random variable.

Definition 2 (Central Moments)**.** For each positive integer n, the nth *central moment* of X is the number $\mathbb{E}[(X - \mathbb{E}[X])^n]$.

The most common central moment is the second one, which is called the variance. This measures the spread of the random variable.

Definition 3 (Variance)**.** The *variance* of X, denoted by $\mathrm{Var}[X]$, is the non-negative number

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

The variance is in the squared units, so the standard deviation defined by

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}[X]}$$

measures the deviation in the original units.

The variance satisfies the following nice properties.

1. Equivalent formula I:
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$
2. Equivalent formula II:
$$\mathrm{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$$
3. Variance of Linear Functions: For any constants $a, b \in \mathbb{R}$,
$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$
4. Variance of Linear Functions II: If X and Y are independent then
$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$
5. Zero Variance: Suppose a random variable X has $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = 0$. This means, X does not "vary" from its mean at all, and is constant with probability 1

   Theorem 1

   $\mathrm{Var}(X) = 0$ *if and only if* $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

Remark 1. The simple formulas to compute the variance is one of the main reasons why we use the squared deviations in the definition of the variance over other measures of deviation. We will see later that squared deviations match our natural ways of measuring distances, so it will allow us to build some geometric intuition of results.

The remainder of this section is devoted to tricks and techniques to compute expected values.

1.1 The Expected Value of Counting Random Variables

Whenever a random variable N takes values in $\{0, 1, 2, \dots, n\}$, we can use the linearity of expectation to compute the expected values in another possibly simpler way. Suppose that N counts the number of events $A_1, A_2, \dots, A_n$ (that are not necessarily independent) that occurred, then

$$N = N(\omega) = \sum_{i=1}^{n} \mathbb{1}(\omega \in A_i) = \sum_{i=1}^{n} \mathbb{1}_{A_i}.$$

Therefore, by the linearity of expectation

$$\mathbb{E}[N] = \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

This trick is especially useful if the joint distribution is tricky to compute, but its marginals are relatively simpler.

1.2 Moment Generating Function

So far we have seen that the distribution of a random variable can always be encoded by the CDF. If we know that the random variable is discrete / continuous, then its distribution can alternatively be encoded by the PMF / PDF. We now show that it is also possible to encode the distribution through another object called the moment generating function provided that the moments of the random variable do not grow too fast.

Definition 4 (Moment Generating Function)**.** The *moment generating function* (MGF) of a random variable X is given by the function

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

provided the expression exists in a neighbourhood of zero, say for $t \in (-a, a)$.

Remark 2. The MGF might not always exist. For example, if the a moments of $\mathbb{E}[X^n]$ of X is infinite, then the MGF will be divergent near zero. In higher level courses, you will encounter characteristic functions or Fourier transforms which will always exist.

1.3 Properties

If X has finite support on $\{x_1, \dots, x_n\}$ with PMF p_X, then

$$M_X(t) = p_X(x_1)e^{tx_1} + \cdots + p_X(x_n)e^{tx_n}$$

so it is easy to read off the PMF given the MGF. We see that even for more complicated distributions the MGF still encodes the distribution of the random variables.

Theorem 2 (*Uniqueness Theorem*)

If X and Y have MGFs $M_X(t)$ and $M_Y(t)$ defined in neighbourhoods of the origin, and $M_X(t) = M_Y(t)$ for all t where they are defined, then

$$X \overset{d}{=} Y.$$

The MGF is also computationally convenient since it can be used to recover the moments and compute the distribution of the sums of independent random variables easily.

1. **Moments:** The MGF encodes all the moments of X. Assuming that $M_X(t)$ is defined in a neighbourhood of $t = 0$,
$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}[X^k] \text{ for all } k \geq 0.$$
This follows from Taylor's theorem and linearity holds even with infinite sums provided that $M_X(t)$ is defined in a neighbourhood of $t = 0$
$$M_X(t) = \mathbb{E}\left[\sum_{j=0}^{\infty} \frac{t^j X^j}{j!}\right] = \sum_{j=0}^{\infty} \frac{t^j \, \mathbb{E}[X^j]}{j!}.$$
Formally taking the kth derivative at 0 gives us the formula (see Problem 1.31 for a careful proof).

2. **Independent Sums:** Suppose that X and Y are independent and have moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. Then the moment generating function of $X + Y$ is
$$M_{X+Y}(t) = \mathbb{E}\left(e^{t(X+Y)}\right) = \mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right) = M_X(t)M_Y(t).$$
This is a useful property because the distribution of the sum of random variables can now be easily computed (which required computing iterated sums).

1.4 Expected Value and Variance of Common Distributions

Distribution	PMF / PDF	Mean	Variance
DUnif(a, b)	$\frac{1}{b-a+1}, \quad x = \{a, a+1, \dots, b\}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Ber(p)	$p^x(1-p)^{1-x}, \quad x \in \{0, 1\}$	p	$p(1-p)$
Bin(n, p)	$\binom{n}{x}p^x(1-p)^{n-x}, \quad x = 0, 1, \dots, n$	np	$np(1-p)$
Geo(p)	$(1-p)^x p, \quad x = 0, 1, 2, \dots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
NegBin(k, p)	$\binom{x+k-1}{x}p^k(1-p)^x, \quad x = 0, 1, 2, \dots$	$\frac{k(1-p)}{p}$	$\frac{k(1-p)}{p^2}$
Poi(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$	λ	λ
Hyp(N, r, n)	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \ \max\{0, n-(N-r)\} \leq x \leq \min\{r, n\}$	$r\frac{n}{N}$	$n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$
Unif(a, b)	$\frac{1}{b-a}, \quad a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exp(θ)	$\frac{1}{\theta}e^{-\frac{x}{\theta}}, \quad x \geq 0$	θ	θ^2
N(μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$	μ	σ^2

1.5 Moment Generating Functions of Common Distributions

1. **Discrete Uniform:** If $X \sim$ DUnif$[a, b]$ (discrete uniform) then
$$M_X(t) = \frac{1}{b-a+1}\sum_{x=a}^{b} e^{tx} \text{ for } t \in \mathbb{R}$$

2. **Binomial:** If $X \sim$ Bin(n, p) then
$$M_X(t) = (pe^t + (1-p))^n \text{ for } t \in \mathbb{R}$$

3. **Negative Binomial:** If $X \sim \text{NegBin}(k, p)$ then

$$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k \text{ for } t < -\ln(1-p)$$

4. **Poisson:** If $X \sim \text{Poi}(\lambda)$ then

$$M_X(t) = e^{\lambda(e^t-1)} \text{ for } t \in \mathbb{R}$$

5. **Continuous Uniform:** If $X \sim \text{Unif}[a, b]$ (continuous uniform) then

$$M_X(t) = \begin{cases} \frac{e^{bt}-e^{at}}{(b-a)t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

6. **Exponential:** If $X \sim \text{Exp}(\theta)$ (waiting time parametrization) then

$$M_X(t) = \frac{1}{1-\theta t} \text{ for } t < \frac{1}{\theta}$$

7. **Normally Distributed:** If $X \sim \text{N}(\mu, \sigma^2)$, then

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \text{ for } t \in \mathbb{R}$$

1.6 Example Problems

Problem 1.1. Consider two games

1. Game 1 pays \$1 with probability $\frac{1}{2}$ and $-\$1$ with probability $\frac{1}{2}$.
2. Game 2 pays \$100 with probability $\frac{1}{2}$ and $-\$100$ with probability $\frac{1}{2}$.

Let X denote the payoff of the first game and Y denote the payoff of the second game. What is the mean and variance of X and Y?

Solution 1.1. This example demonstrates the necessity of the variance. Both games have the same expected value, but their spread is very different, so the expected value is not enough to capture all the interesting behaviors we care about. We have

$$\mathbb{E}[X] = +1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0 \qquad \text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 1 - 0^2 = 1,$$

and

$$\mathbb{E}[Y] = +100 \cdot \frac{1}{2} + (-100) \cdot \frac{1}{2} = 0 \qquad \text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = 100^2 - 0^2 = 10000.$$

Both games are fair game, but the variance of game 2 is much higher than the variance of game 1.

Problem 1.2. Consider the random variables

- X is a r.v. representing the outcome of one fair 6-sided die roll
- Y is a r.v. representing the number of phone calls over 1 minute at Lenovo call centre, with the rate of 3.5 calls per minute

Compute the mean and variance of X and Y.

Solution 1.2. We have that $X \sim \text{DUnif}[1,6]$ and $Y \sim \text{Poi}(3.5)$. Using the formulas for mean and variance, we have

$$\mathbb{E}[X] = 3.5,\ \text{Var}(X) = \frac{6^2-1}{12} \approx 2.9$$

while

$$\mathbb{E}[Y] = 3.5,\ \text{Var}(Y) = 3.5.$$

This makes intuitive sense because both X and Y have the same mean, but the fact that Y can take values on $\mathbb{N}$ while X can only take values on $\{1,2,\dots,6\}$, so Y should have larger variance even though both random variables have the same mean.

Problem 1.3. Suppose that X has variance $\text{Var}(X) = 2$. Compute the variance of Y, where $Y = -2X + 3$.

Solution 1.3. By the variance of linear maps,

$$\text{Var}(Y) = \text{Var}(-2X+3) = (-2)^2\,\text{Var}(X) = 8.$$

Problem 1.4. Let $n \geq 1$. Show that

$$\Gamma(n) = \int_0^\infty y^{n-1}e^{-y}dy = (n-1)!.$$

Solution 1.4. This follows from repeated integration by parts. Let $n \geq 1$, we have

$$\Gamma(n) = \int_0^\infty y^{n-1}e^{-y}dy.$$

Integrating by parts we see that

$$\Gamma(n) = \int_0^\infty y^{n-1}e^{-y}dy = -y^{n-1}e^{-y}\Big|_0^\infty + (n-1)\int_0^\infty y^{n-2}e^{-y}dy = (n-1)\Gamma(n-1).$$

Repeating this inductively implies that

$$\Gamma(n) = (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \cdots = (n-1)(n-2)\cdots 3\cdot 2\cdot \Gamma(1) = (n-1)!$$

since $\Gamma(1) = 1$ (using the fact that the integral of the PDF of $X \sim \text{Exp}(1)$ is 1).

Remark 3. Notice that the integration by parts computation holds for $\alpha > 1$ which are not necessarily integer valued to conclude that $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. The problem is that the inductive step might not lead to a nice number since $\Gamma(x)$ for $x < 1$ doesn't necessarily simplify.

Problem 1.5. A continuous random variable X is said to have a Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ if it has PDF

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Use the properties of the Gamma function, namely $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ for $\alpha > 1$, to obtain $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution 1.5. We can solve this problem without even knowing what the $\Gamma(\alpha)$ function is. We use a trick that we have used many times, which involves rewriting the expression as an integral of the PDF which sums to 1.

Expected Value: By definition,

$$\begin{aligned}\mathbb{E}[X] = \int_0^\infty x \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha} e^{-\frac{x}{\beta}}\, dx \\ &= \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^{\alpha} e^{-\frac{x}{\beta}}\, dx}_{=1 \text{ integral of the PDF of } \Gamma(\alpha+1,\beta) \text{ r.v.}}\end{aligned}$$

where we multiplied and divided by $\beta\Gamma(\alpha+1)$ to match the normalization terms. Using the recursive property of the Γ function, we see that

$$\mathbb{E}[X] = \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\beta\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \beta\alpha.$$

Variance: A similar computation as above implies that

$$\begin{aligned}\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha+1} e^{-\frac{x}{\beta}}\, dx \\ &= \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha+2)\beta^{\alpha+2}} x^{\alpha+1} e^{-\frac{x}{\beta}}\, dx}_{=1 \text{ integral of the PDF of } \Gamma(\alpha+2,\beta) \text{ r.v.}} \\ &= \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)}.\end{aligned}$$

Next, using the recursive properties of the Γ function,

$$\mathbb{E}[X^2] = \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} = \frac{\beta^2(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\beta^2(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \beta^2(\alpha+1)\alpha.$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \beta^2(\alpha+1)\alpha - \beta^2\alpha^2 = \beta^2\alpha.$$

Remark 4. Notice that then $\alpha = 1$, then the PDF is simply the exponential distribution. In fact, the relation is even deeper and it can be shown that the sum of n independent $\mathrm{Exp}(\beta)$ distributions has a $\Gamma(n,\beta)$ distribution.

Problem 1.6. *n passengers board a plane with n seats, where $n > 1$. Despite every passenger having an assigned seat, when they board the plane they sit in one of the remaining available seats at random. Show that the mean and variance of the number of people sitting in the correct seat once everyone is on board are both 1 (independent of the number n of passengers, weirdly enough).*

Solution 1.6. This is called the matching problem. Let N denote the number of people sitting in the correct of seat once everyone is on board, and let A_i be the event that the ith passenger is in the correct seat. We have

$$\mathbb{1}_{A_i} = \begin{cases} 1 & \text{the } i\text{th passenger is in the correct seat} \\ 0 & \text{the } i\text{th passenger is not in the correct seat .}\end{cases}$$

Clearly, $N = \sum_{i=1}^n \mathbb{1}_{A_i}$. We can now compute the mean and variance.

Expected Value: By linearity of expectation

$$\mathbb{E}[N] = \sum_{i=1}^n \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^n \mathbb{P}(A_i).$$

By symmetry, we have that the probability that the ith passenger is in the correct seat is

$$\mathbb{P}(A_i) = \frac{1}{n}$$

since the seat the ith passenger sits in is uniform over the n possible seats. Therefore,

$$\mathbb{E}[N] = \sum_{i=1}^n \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{i=1}^n \frac{1}{n} = 1.$$

Variance: By the linearity of expectation

$$\mathbb{E}[N^2] = \mathbb{E}\left[\left(\sum_{i=1}^n \mathbb{1}_{A_i}\right)^2\right] = \sum_{i,j=1}^n \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}].$$

We have two cases

1. $i = j$: Suppose that $i = j$. Since $\mathbb{1}(A_i)\mathbb{1}(A_i) = 1$ if an only if A_i happens, so we have

   $$\mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_i}] = \mathbb{E}[\mathbb{1}_{A_i}] = \mathbb{P}(A_i) = \frac{1}{n}$$

   as we computed before.

2. $i \neq j$: Suppose that $i \neq j$. Since $\mathbb{1}_{A_i}\mathbb{1}_{A_j} = 1$ if an only if A_i and A_j happens

   $$\mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \mathbb{P}(A_i \cap A_j) = \frac{1}{n(n-1)}.$$

   Note that the events A_i and A_j are not independent, so we can't simply multiply the probabilities. Instead, we can use the fact that sets the i and j passengers sit in are uniform over the $n(n-1)$ possible seats for two passengers.

Since there are $n(n-1)$ ways to pick indices $i \neq j$ and n ways to pick indices $i = j$, we have

$$\mathbb{E}[N^2] = \sum_{i,j=1}^n \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \sum_{i=j} \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_i}] + \sum_{i\neq j} \mathbb{E}[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \frac{n}{n} + \frac{n(n-1)}{n(n-1)} = 2.$$

Therefore,

$$\mathrm{Var}(N) = \mathbb{E}[N^2] - (\mathbb{E}[N])^2 = 2 - 1 = 1.$$

Remark 5. Notice that the events $A_1, \dots A_n$ are clearly not independent. For example, if $A_1, \dots, A_{n-1}$ were to happen then A_n must be true too since the seat left is the one assigned to the last passenger. The linearity of expectation allowed us to decompose the random variable into a sum of possibly dependent events. However, by symmetry we only needed to compute the probability of a single event A_1 in isolation without worrying about the other events $A_2, \dots, A_n$.

Remark 6. Instead of using the uniform distribution and symmetry, we could argue that

$$\mathbb{P}(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

since there are $(n-1)!$ seating patterns where the ith passenger is in the right seat and $n!$ total seating patterns (all of which are equally likely). Likewise, we have

$$\mathbb{P}(A_i \cap A_j) = \frac{1}{n(n-1)} = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

since there are $(n-2)!$ seating patterns where the ith and jth passenger is in the right seat and $n!$ total seating patterns (all of which are equally likely).

Yet another way to compute the probability is to argue sequentially using the chain rule,

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i \mid A_j)\,\mathbb{P}(A_j) = \frac{1}{n-1} \cdot \frac{1}{n} = \frac{1}{n(n-1)},$$

since the probability the jth passenger sits in the right seat is $\frac{1}{n}$ and the probability the ith passenger sits in the right seat given that the jth passenger is in the right seat is $\frac{1}{n-1}$ since the jth passenger is already in the correct seat so there are $n-1$ seats left.

Problem 1.7. Suppose that X has MGF

$$M_X(t) = 0.3 + 0.2e^t + 0.5e^{2t}$$

What is the PMF of X?

Solution 1.7. We have

$$M_X(t) = 0.3 + 0.2e^t + 0.5e^{2t} = 0.3e^{0\cdot t} + 0.2e^{1\cdot t} + 0.5e^{2\cdot t}.$$

Then reverse engineering the PMF from the MGF implies that

$$p_X(0) = \mathbb{P}(X=0) = 0.3, \qquad p_X(1) = \mathbb{P}(X=1) = 0.2 \qquad p_X(2) = \mathbb{P}(X=2) = 0.5.$$

It is clear that at least for finitely supported PMF, the MGF is simply another way of writing encoding the distribution.

Problem 1.8. Use the MGFs to show that if $X \sim \mathrm{Poi}(\lambda)$, then $\mathbb{E}[X] = \lambda$ and $\mathrm{Var}(X) = \lambda$

Solution 1.8. The MGF is

$$M_X(t) = e^{\lambda(e^t-1)}.$$

Taking derivatives, we have

$$M_X'(t) = e^{\lambda(e^t-1)}\lambda e^t \Rightarrow \mathbb{E}(X) = M_X'(0) = \lambda.$$

To compute the variance, we take the second derivative

$$M_X''(t) = e^{\lambda(e^t-1)}\lambda e^t + e^{\lambda(e^t-1)}(\lambda e^t)^2 \Rightarrow \mathbb{E}(X^2) = M_X''(0) = \lambda^2 + \lambda$$

and conclude

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

Problem 1.9. Use MGFs to show that if $X \sim \mathrm{Poi}(\lambda)$ and $Y \sim \mathrm{Poi}(\mu)$ are independent, then $X+Y \sim \mathrm{Poi}(\lambda+\mu)$.

Solution 1.9. We know that $M_X(t) = e^{\lambda(e^t-1)}$ and $M_Y(t) = e^{\mu(e^t-1)}$. Since X and Y are independent, the MGF of $X+Y$ is

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1)}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

which we recognize as the MGF of a $\mathrm{Poi}(\lambda+\mu)$ random variables. By uniqueness of the MGF, we conclude $X+Y \sim \mathrm{Poi}(\lambda+\mu)$.

Problem 1.10. Use MGFs to show that if $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ are independent, then $X+Y \sim \mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Solution 1.10. We know that $M_X(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}}$ and $M_Y(t) = e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}}$. Since X and Y are independent, the MGF of $X+Y$ is

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} = e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

which we recognize as the MGF of a $\mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$ random variables. By uniqueness of the MGF, we conclude $X+Y \sim \mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Problem 1.11. Use MGFs to show that for large n and small p such that $\mu = np$, we can approximate the $\mathrm{Bin}(n,p)$ distribution with a $\mathrm{Poi}(\mu)$ distribution.

Solution 1.11. By the uniqueness theorem, we can prove this by showing that the MGF of $X \sim \mathrm{Bin}(n,p)$ converges to the MGF of $Y \sim \mathrm{Poi}(\mu)$ as $n \to \infty$ where $\mu = np \Leftrightarrow p = \frac{\mu}{n}$. Indeed,

$$M_X(t) = (pe^t + 1 - p)^n = (1 + p(e^t-1))^n = \Big(1 + \frac{\mu}{n}(e^t-1)\Big)^n.$$

Since for every x

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

we have

$$\lim_{n\to\infty} M_X(t) = \lim_{n\to\infty} \Big(1 + \frac{\mu}{n}(e^t-1)\Big)^n = e^{\mu(e^t-1)} = M_Y(t),$$

Problem 1.12. Let $Z \sim \mathrm{N}(0,1)$. What is the distribution of $X = \sigma Z + \mu$?

Solution 1.12. This result is an alternative proof of the de-standardization argument. If $Z \sim \mathrm{N}(0,1)$, then

$$M_Z(t) = e^{\frac{t^2}{2}}.$$

If we define $X = \sigma Z + \mu$, then the MGF of X satisfies

$$M_X(t) = M_{\sigma Z+\mu}(t) = \mathbb{E}[e^{t\sigma Z + t\mu}] = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{\frac{\sigma^2}{2} t^2}$$

which we recognize as the MGF of $X \sim \mathrm{N}(\mu, \sigma^2)$.

Problem 1.13. Let $X \sim \text{Exp}(1)$. What is the distribution of $Y = \theta X$?

Solution 1.13. If $X \sim \text{Exp}(1)$, then
$$M_X(t) = \frac{1}{1-t}.$$
If we define $Y = \theta X$, then the MGF of Y satisfies
$$M_Y(t) = M_{\theta X}(t) = \mathbb{E}[e^{t\theta X}] = M_X(\theta t) = \frac{1}{1-\theta t}$$
which we recognize as the MGF of $Y \sim \text{Exp}(\theta)$.

1.7 Proofs of Key Results

Problem 1.14. Show that
$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2.$$

Solution 1.14. To simplify notation, we define $\mathbb{E}[X] = \mu$. Then by the linearity of expectation,
$$\text{Var}(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}[X^2] - 2\mu\,\mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$
To conclude the second equality, notice that
$$\mathbb{E}[X(X-1)] = \mathbb{E}[X^2 - X] = \mathbb{E}[X^2] - \mathbb{E}[X] \implies \mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X]$$
so substituting this into the formula above implies
$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2.$$

Problem 1.15. Show that for any constants a and b,
$$\text{Var}(aX+b) = a^2\,\text{Var}(X).$$

Solution 1.15. We can use the definition of the variance. Let $Y = aX + b$,
$$\begin{aligned}
\text{Var}(aX+b) = \text{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] &= \mathbb{E}[(aX + b - \mathbb{E}[aX+b])^2] \\
\mathbb{E}[aX+b] = a\,\mathbb{E}[X] + b \qquad &= \mathbb{E}[(aX - a\,\mathbb{E}[X])^2] \\
&= a^2\,\mathbb{E}[(X - \mathbb{E}[X])^2] = a^2\,\text{Var}(X).
\end{aligned}$$

Remark 7. This makes intuitive sense because shifting a random variable by b does not change the spread of the random variables. However, scaling the random variable by a will change the spread by a factor of a^2 since we are measuring the squared deviations, so the scaling factor is squared.

Problem 1.16. Prove Theorem 1.

Solution 1.16. Let $\mathbb{E}[X] = \mu$.

($\Longrightarrow$) Suppose $\mathbb{P}(X = \mu) = 1$. In other words, $p_X(\mu) = 1$ and there are no other non-zero values of the PMF, so by definition of the variance,

$$\mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu^2\,\mathbb{P}(X=\mu) - (\mu\,\mathbb{P}(X=\mu))^2 = 0.$$

($\Longleftarrow$) Conversely, suppose $Var(X) = 0$. Then, again by definition of the variance,

$$0 = \mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \sum_{\text{all } x} \underbrace{(x-\mu)^2}_{\geq 0}\underbrace{\mathbb{P}(X=x)}_{=p_X(x)\geq 0}.$$

Suppose for the sake of contradiction that there exists a $\nu \neq \mu$ such that $\mathbb{P}(X = \nu) > 0$. In this case, we have

$$\mathrm{Var}(X) \geq (\nu-\mu)^2\,\mathbb{P}(X=\nu) > 0$$

which contradicts the fact that $\mathrm{Var}(X) = 0$. Therefore, we must have $\mathbb{P}(X = \mu) = 1$.

Problem 1.17. If $X \sim \mathrm{DUnif}[a, b]$ then $\mathbb{E}[X] = \frac{a+b}{2}$.

Solution 1.17. If $X \sim \mathrm{DUnif}[a, b]$ then the sum of positive integers implies

$$\begin{aligned}\mathbb{E}[X] = \sum_{x=a}^{b} \frac{x}{b-a+1} = \frac{1}{b-a+1}\left(\sum_{x=1}^{b} x - \sum_{x=1}^{a-1} x\right) &= \frac{1}{b-a+1}\left(\frac{b(b+1)}{2} - \frac{a(a-1)}{2}\right)\\ &= \frac{1}{b-a+1}\left(\frac{b^2-a^2+(b+a)}{2}\right) = \frac{a+b}{2}.\end{aligned}$$

Alternative Solution: If $X \sim \mathrm{DUnif}[0, n]$ then the sum of positive integers implies

$$\mathbb{E}[X] = \sum_{x=0}^{n} \frac{x}{n+1} = \frac{n(n+1)}{2(n+1)} = \frac{n}{2}.$$

Notice that $X \sim \mathrm{DUnif}[a, b] \sim a + \mathrm{DUnif}[0, b-a]$. Therefore, if we let $Y \sim \mathrm{DUnif}[0, b-a]$, then $X = a + Y$ so linearity implies that

$$\mathbb{E}[X] = \mathbb{E}[a+Y] = a + \mathbb{E}[Y] = a + \frac{b-a}{2} = \frac{a+b}{2}.$$

Problem 1.18. If $X \sim \mathrm{Hyp}(N, r, n)$, then $\mathbb{E}[X] = r\frac{n}{N}$.

Solution 1.18. This proof uses a trick called the linearity of expectation. To simplify notation, suppose that we have r blue balls and $N - r$ red balls, then $X \sim \mathrm{Hyp}(N, r, n)$ denotes the number of blue balls we drew from a sample of n balls without replacement.

We label the blue balls $1, \ldots, r$ and let A_i denote the event that the blue ball labeled i was drawn. Consider the random variable

$$\mathbb{1}_{A_i} = \begin{cases} 1 & \text{if we drew the blue ball labeled } i \\ 0 & \text{if we did not draw the blue ball labeled } i. \end{cases}$$

If $X \sim \text{Hyp}(N, r, n)$ then $X = \sum_{i=1}^{r} \mathbb{1}(A_i)$ which is the total number of blue balls that we drew. We have by the linearity of expectation that

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{r} \mathbb{1}_{A_i}\right] = \sum_{i=1}^{r} \mathbb{E}[\mathbb{1}_{A_i}].$$

Next, for any i, we have

$$\mathbb{E}[\mathbb{1}_{A_i}] = 1 \cdot \mathbb{P}(A_1) + 0 \cdot (1 - \mathbb{P}(A_1)) = \mathbb{P}(\text{we drew the blue ball labeled } i).$$

By symmetry (we are not more likely to draw a particular ball over another one), for all $i \le r$

$$\mathbb{P}(A_i) = \mathbb{P}(A_1) = \mathbb{P}(\text{we drew the blue ball labeled 1}) = \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

so

$$\mathbb{E}[X] = \sum_{i=1}^{r} \mathbb{E}[\mathbb{1}_{A_i}] = r\,\mathbb{P}(A_1) = \frac{rn}{N}.$$

Problem 1.19. If $X \sim \text{Bin}(n, p)$ then $\mathbb{E}[X] = np$.

Solution 1.19. If $X \sim \text{Bin}(n, p)$ then $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ so

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \cdot \frac{n!}{(n-x)!x!} \cdot p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n \cdot (n-1)!}{((n-1)-(x-1))!(x-1)!} p \cdot p^{x-1} (1-p)^{n-1-(x-1)} && \text{add and subtract 1} \\
&= np \sum_{y=0}^{n-1} \frac{(n-1)!}{((n-1)-y)!y!} \cdot p^y (1-p)^{(n-1)-y} && \text{re-index sum} \\
&= np \underbrace{\sum_{y=0}^{n-1} \binom{n-1}{y} \cdot p^y (1-p)^{(n-1)-y}}_{=1 \text{sum of PMF of Bin}(n-1,p)} \\
&= np.
\end{aligned}$$

Alternative Solution: If $X_i \sim \text{Ber}(p)$ then

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p.$$

Now suppose that $X \sim \text{Bin}(n, p)$. Since $X = X_1 + \cdots + X_n$ where X_i are independent and $X_i \sim \text{Ber}(p)$ (the number of successes in n trials is equal to the sum of n successful trials), linearity implies that

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \dots X_n] = n\,\mathbb{E}[X_1] = np.$$

Problem 1.20. If $X \sim \text{NegBin}(k, p)$, show that $\mathbb{E}[X] = \frac{k(1-p)}{p}$.

Solution 1.20. We first consider the geometric random variable. We will use many times throughout this derivation the identity

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

for $|q| < 1$. If $X \sim \text{Geo}(p) \sim \text{NegBin}(1, p)$ then $p_X(x) = p(1-p)^x$ so

$$\begin{aligned}
\mathbb{E}[X] = \sum_{x=0}^{\infty} xp(1-p)^x &= p \sum_{x=1}^{\infty} \sum_{k=1}^{x} (1-p)^x && 0 \cdot e^p(1-p)^0 = 0, \sum_{k=1}^{x} = x \\
&= p \sum_{k=1}^{\infty} \sum_{x=k}^{\infty} (1-p)^x && 1 \le k \le x < \infty \\
&= p \sum_{k=1}^{\infty} (1-p)^k \sum_{x=0}^{\infty} (1-p)^x && \sum_{x=k}^{\infty} (1-p)^x = (1-p)^k \sum_{x=0}^{\infty} (1-p)^x \\
&= p \sum_{k=1}^{\infty} \frac{(1-p)^k}{1-(1-p)} && \sum_{x=0}^{\infty} (1-p)^x = \frac{1}{1-(1-p)} \\
&= (1-p) \sum_{k=0}^{\infty} (1-p)^k && \sum_{x=1}^{\infty} (1-p)^x = (1-p) \sum_{x=0}^{\infty} (1-p)^x \\
&= \frac{1-p}{p}. && \sum_{x=0}^{\infty} (1-p)^x = \frac{1}{1-(1-p)}
\end{aligned}$$

Now suppose that $X \sim \text{NegBin}(k, p)$. For $1 \le i \le k$, let X_i denote the number of fails between the $(i-1)$st success and the ith success. Since X_i counts the number of fails until the next success, we have $X_i \sim \text{Geo}(p)$ for all i. By definition, $X = X_1 + \cdots + X_k$ since the total fails until k successes is equal to the sum of the number fails between successes, linearity implies that

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \dots X_k] = k\,\mathbb{E}[X_1] = \frac{k(1-p)}{p}.$$

Problem 1.21. If $X \sim \text{Poi}(\lambda)$, show that $\mathbb{E}[X] = \lambda$.

Solution 1.21. If $X \sim \text{Poi}(\lambda)$ then $p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ so

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{\infty} x \cdot e^{-\lambda} \frac{\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x \cdot e^{-\lambda} \frac{\lambda^x}{x!} && 0 \cdot e^{-\lambda} \frac{\lambda^0}{0!} = 0 \\
&= \lambda \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda \underbrace{\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!}}_{=1 \text{ sum of PMF of Poi}(\lambda)} && \text{re-index sum} \\
&= \lambda.
\end{aligned}$$

Problem 1.22. If $X \sim \mathrm{DUnif}[a,b]$ then $\mathrm{Var}(X) = \frac{(b-a+1)^2-1}{12}$.

Solution 1.22. If $X \sim \mathrm{DUnif}[0,n]$ then the sum of positive integers implies

$$\mathbb{E}[X] = \sum_{x=0}^{n} \frac{x}{n+1} = \frac{n(n+1)}{2(n+1)} = \frac{n}{2}, \quad \mathbb{E}[X^2] = \sum_{x=0}^{n} \frac{x^2}{n+1} = \frac{n(n+1)(2n+1)}{6(n+1)} = \frac{n(2n+1)}{6}$$

so

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

Notice that $X \sim \mathrm{DUnif}[a,b] \sim a + \mathrm{DUnif}[0,b-a]$. Therefore, if we let $Y \sim \mathrm{DUnif}[0,b-a]$, then $X = a+Y$ so the variance of a linear function implies that

$$\mathrm{Var}(X) = \mathrm{Var}(a+Y) = \mathrm{Var}(Y) = \frac{(b-a+1)^2-1}{12}.$$

Problem 1.23. If $X \sim \mathrm{Hyp}(N,r,n)$, then $\mathrm{Var}(X) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$.

Solution 1.23. This proof uses a trick called the linearity of expectation. To simplify notation, suppose that we have r blue balls and $N-r$ red balls, then $X \sim \mathrm{Hyp}(N,r,n)$ denotes the number of blue balls we drew from a sample of n balls without replacement. We first compute $\mathbb{E}[X(X-1)]$. It suffices to compute

$$\mathbb{E}\left[\binom{X}{2}\right] = \mathbb{E}\left[\frac{X(X-1)}{2}\right]$$

which denotes the expected value of the number of pairs of blue balls we drew.

We label the successful balls $1,\dots,r$ and let A_{ij} denote the event that the pair of balls labeled i and j was drawn where. Consider the random variable

$$\mathbb{1}_{A_{ij}} = \begin{cases} 1 & \text{if we drew the pair of blue balls } i \text{ and } j \\ 0 & \text{if we did not draw the pair of blue balls } i \text{ and } j. \end{cases}$$

If $X \sim \mathrm{Hyp}(N,r,n)$ then $\binom{X}{2} = \sum_{1\le i<j\le r} \mathbb{1}_{A_{ij}}$ which is the total number of pairs of blue balls that we drew. We have by the linearity of expectation that

$$\mathbb{E}\left[\binom{X}{2}\right] = \mathbb{E}\left[\sum_{1\le i<j\le r} \mathbb{1}_{A_{ij}}\right] = \sum_{1\le i<j\le r} \mathbb{E}[\mathbb{1}_{A_{ij}}].$$

Next, for any i,j, we have

$$\mathbb{E}[\mathbb{1}_{A_{ij}}] = 1\,\mathbb{P}(A_{ij}) + 0\cdot(1-\mathbb{P}(A_{ij})) = \mathbb{P}(A_{ij}).$$

By symmetry (we are not more likely to draw a particular ball over another one), for all $1 \le i < j \le r$

$$\mathbb{P}(A_{ij}) = \mathbb{P}(A_{12}) = \mathbb{P}(\text{ we drew the pair of blue balls } i \text{ and } j\ ) = \frac{\binom{2}{2}\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

so

$$\mathbb{E}\left[\binom{X}{2}\right] = \sum_{1\le i<j\le r} \mathbb{E}[\mathbb{1}(A_{ij})] = \frac{r(r-1)}{2}\,\mathbb{P}(A_{12}) = \frac{r(r-1)}{2}\frac{n(n-1)}{N(N-1)}.$$

Multiplying by two implies that

$$\mathbb{E}[X(X-1)] = 2\,\mathbb{E}\left[\binom{X}{2}\right] = 2\,\mathbb{E}\left[\frac{X(X-1)}{2}\right] = \frac{r(r-1)n(n-1)}{N(N-1)}.$$

Since we know that $\mathbb{E}[X] = r\frac{n}{N}$ we have

$$\operatorname{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \frac{r(r-1)n(n-1)}{N(N-1)} + \frac{rn}{N} - \frac{r^2n^2}{N^2} = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Remark 8. We can set $p = \frac{r}{N}$ which denotes the probability of a successful draw. If $X \sim \operatorname{Hyp}(N, r, n)$ and $Y \sim \operatorname{Bin}(n, p)$ then

$$\mathbb{E}[X] = np \qquad \operatorname{Var}(X) = np(1-p)\left(\frac{N-n}{N-1}\right).$$

and

$$\mathbb{E}[Y] = np \qquad \operatorname{Var}(Y) = np(1-p).$$

The hypergeometric and binomial random variable have the same mean, and they have the same variance except for a factor $\frac{N-n}{N-1} \leq 1$. The variance of the hypergeometric is slightly less because sampling without replacement reduces the "spread" since our sample space shrinks with each draw.

When $n = 1$ then the factor $\frac{N-n}{N-1} = 1$, so the variance of a hypergeometric and binomial random variables are the same, since sampling one object with or without replacement is the same. When $N \to \infty$ then the factor $\frac{N-n}{N-1} \to 1$, which is again consistent because sampling with or without replacement from a large population is essentially the same.

Problem 1.24. If $X \sim \operatorname{Bin}(n, p)$ then $\operatorname{Var}(X) = np(1-p)$.

Solution 1.24. If $X \sim \operatorname{Bin}(n, p)$ then $p_X(x) = \binom{n}{x}p^x(1-p)^{n-x}$. We use the formula

$$\operatorname{Var}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2$$

and note we already know $\mathbb{E}(X) = np$. By definition

$$\begin{aligned}
\mathbb{E}(X(X-1)) &= \sum_{x=0}^{n} x(x-1)\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-2-(x-2))!(x-2)!}p^{x-2}(1-p)^{n-2-(x-2)} \\
&= n(n-1)p^2 \underbrace{\sum_{y=0}^{n-2} \frac{(n-2)!}{(n-2-y)!y!}p^y(1-p)^{n-2-y}}_{=1 \text{ sum of PMF of } (n-2,p)} \\
&= n(n-1)p^2.
\end{aligned}$$

Then

$$\operatorname{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Alternative Solution: If $X \sim \operatorname{Bern}(p)$ then by definition

$$\operatorname{Var}(X) = \mathbb{E}[(X-p)^2] = (1-p)^2p + (-p)^2(1-p) = p(1-p).$$

Now suppose that $X \sim \operatorname{Bin}(n, p)$. Since $X = X_1 + \cdots + X_n$ where X_i are independent and $X_i \sim \operatorname{Ber}(p)$ (the number of successes in n trials is equal to the sum of n successful trials), linearity implies that

$$\operatorname{Var}(X) = \operatorname{Var}(X_1 + \ldots X_n) = n\operatorname{Var}(X_1) = np(1-p).$$

Problem 1.25. If $X \sim \text{NegBin}(k,p)$, show that $\text{Var}(X) = \frac{k(1-p)}{p^2}$.

Solution 1.25. We first consider the geometric random variable. We will use many times throughout this derivation the identity

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

for $|q| < 1$. From this identity, we can recover higher order versions of this by differentiating the power series term by term with respect to q, that is

$$\frac{d}{dq}\sum_{k=0}^{\infty} q^k = \frac{d}{dq}\frac{1}{1-q} \implies \sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}$$

and

$$\frac{d^2}{dq^2}\sum_{k=0}^{\infty} q^k = \frac{d^2}{dq^2}\frac{1}{1-q} \implies \sum_{k=2}^{\infty} k(k-1)q^{k-2} = \frac{2}{(1-q)^3}$$

We will use these identities to give a different derivation of the first and second moments of a geometric random variable.

If $X \sim \text{Geo}(p) \sim \text{NegBin}(1,p)$ then $p_X(x) = p(1-p)^x$ so the first derivative identity implies that

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} xp(1-p)^x = \sum_{x=1}^{\infty} xp(1-p)^x = p(1-p)\sum_{x=1}^{\infty} x(1-p)^{x-1} = \frac{p(1-p)}{(1-(1-p))^2} = \frac{(1-p)}{p}.$$

Likewise, the second derivative identity implies that

$$\mathbb{E}[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)p(1-p)^x = p(1-p)^2\sum_{x=2}^{\infty} x(x-1)(1-p)^{x-2} = \frac{2p(1-p)^2}{(1-(1-p))^3} = \frac{2(1-p)^2}{p^2}.$$

Therefore,

$$\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \frac{2(1-p)^2}{p^2} + \frac{(1-p)}{p} - \frac{(1-p)^2}{p^2} = \frac{(1-p)}{p^2}.$$

Now suppose that $X \sim \text{NegBin}(k,p)$. For $1 \le i \le k$, let X_i denote the number of fails between the $(i-1)$st success and the ith success. Since X_i counts the number of fails until the next success, we have $X_i \sim \text{Geo}(p)$ for all i and the X_i are independent. By definition, $X = X_1 + \cdots + X_k$ since the total fails until k successes is equal to the sum of the number fails between successes, linearity implies that

$$\text{Var}(X) = \text{Var}(X_1 + \ldots X_k) = k\,\text{Var}(X_1) = \frac{k(1-p)}{p^2}.$$

Problem 1.26. If $X \sim \text{Poi}(\lambda)$, show that $\text{Var}(X) = \lambda$.

Solution 1.26. If $X \sim \text{Poi}(\lambda)$ then $p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$. We use the formula

$$\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$$

and note we already know $\mathbb{E}(X) = \lambda$. By definition

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1) \cdot e^{-\lambda} \frac{\lambda^x}{x!} \\
&= \sum_{x=2}^{\infty} x(x-1) \cdot e^{-\lambda} \frac{\lambda^x}{x!} && 0 \cdot e^{-\lambda} \frac{\lambda^0}{0!} = 0 \\
&= \lambda^2 \sum_{x=2}^{\infty} e^{-\lambda} \frac{\lambda^{x-2}}{(x-2)!} \\
&= \lambda^2 \underbrace{\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!}}_{=1 \text{ sum of PMF of Poi}(\lambda)} && \text{re-index sum} \\
&= \lambda^2.
\end{aligned}$$

Then

$$\mathrm{Var}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

Problem 1.27. Compute the mean and variance of $X \sim \mathrm{Unif}(a, b)$.

Solution 1.27. We can directly compute

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x)\, dx = \int_a^b x \cdot \frac{1}{b-a}\, dx \\
&= \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}.
\end{aligned}$$

To compute the variance, we have

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_a^b x^2 \frac{1}{b-a}\, dx \\
&= \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.
\end{aligned}$$

so after some algebra,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{b^2 + ab + a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

Problem 1.28. Compute the mean and variance of $X \sim \mathrm{Exp}(\theta)$.

Solution 1.28. We use the change of variable $x = y\theta$ with $dx = \theta dy$

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^{\infty} x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx \overset{(x = y\theta)}{=} \int_0^{\infty} y e^{-y} \theta dy \\
&= \theta \underbrace{\int_0^{\infty} y e^{-y} dy}_{=\Gamma(2)} = \theta \Gamma(2) = \theta \cdot (1!) = \theta.
\end{aligned}$$

To compute the variance, we have

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx \overset{(x=y\theta)}{=} \int_0^\infty \theta y^2 e^{-y} \theta dy \\ &= \theta^2 \underbrace{\int_0^\infty y^{3-1} e^{-y} dy}_{=\Gamma(3)} = \theta^2 \Gamma(3) = \theta^2 \cdot (2!) = 2\theta^2.\end{aligned}$$

so

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

Alternative Solution: We can also integrate by parts directly without using the Gamma function. Starting with

$$\mathbb{E}[X] = \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

we have by integrating by parts

$$\mathbb{E}[X] = \int_0^\infty x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \frac{x}{\theta} \cdot (-\theta e^{-\frac{x}{\theta}}) - \frac{1}{\theta} \theta^2 e^{-\frac{x}{\theta}} \Big|_0^\infty = \theta.$$

Next, to compute

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx$$

we have by integrating by parts

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}}\, dx = \frac{x^2}{\theta} \cdot (-\theta e^{-\frac{x}{\theta}}) - \frac{2x}{\theta} \cdot \theta^2 e^{-\frac{x}{\theta}} + \frac{2}{\theta} \cdot (-\theta^3 e^{-\frac{x}{\theta}}) \Big|_0^\infty = 2\theta^2.$$

Problem 1.29. If $X \sim \mathrm{N}(\mu, \sigma^2)$ show that

$$\mathbb{E}[X] = \mu \qquad \mathrm{Var}(X) = \sigma^2.$$

Solution 1.29. We first do the computations when $Z \sim \mathrm{N}(0,1)$. We have

$$\mathbb{E}[Z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty x e^{\frac{-x^2}{2}} = 0$$

because $xe^{\frac{-x^2}{2}}$ is an odd function. Next, to compute the second moment, we can integrate by parts

$$\mathbb{E}[Z^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty x^2 e^{\frac{-x^2}{2}} = -\frac{x}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \Big|_{-\infty}^\infty + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{\frac{-x^2}{2}}\, dx = 1$$

where the boundary term vanishes because $e^{\frac{-x^2}{2}}$ goes to zero faster than x and the second term is the integral of the PDF of a standard normal. Hence we have shown that

$$\mathbb{E}[Z] = 0 \qquad \mathrm{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = 1.$$

To get the mean and variance of $X \sim \mathrm{N}(\mu, \sigma^2)$, we can use the standardization trick to write $X \overset{d}{=} \sigma Z + \mu$. Therefore,

$$\mathbb{E}[X] = \mathbb{E}[\sigma Z + \mu] = \sigma\, \mathbb{E}[Z] + \mu = \mu$$

and

$$\mathrm{Var}(X) = \mathrm{Var}(\sigma Z + \mu) = \sigma^2\, \mathrm{Var}(Z) = \sigma^2.$$

Problem 1.30. Prove the following formulas for the MGFs of common distributions

1. **Poisson:** If $X \sim \mathrm{Poi}(\lambda)$ then
$$M_X(t) = e^{\lambda(e^t-1)} \text{ for } t \in \mathbb{R}$$
2. **Normal Distribution:** If $X \sim \mathrm{N}(\mu, \sigma^2)$, then
$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \text{ for } t \in \mathbb{R}$$
3. **Negative Binomial:** If $X \sim \mathrm{NegBin}(k, p)$ then
$$M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k \text{ for } t < -\ln(1-p)$$

Solution 1.30. The computation of all the MGFs use similar tricks we have seen before (summation formulas, reducing to sums of PMF/integral PDF, etc)

Poisson: The MGF is computed using the exponential sum
$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t-1)}.$$

Normal Distribution: The MGF is computed by completing the square. We have
$$M_X(t) = \mathbb{E}(e^{tX}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2(\mu+t\sigma^2)x+\mu^2}{2\sigma^2}}\, dx.$$
We want to rewrite the integral in terms of the integral of a PDF, so we complete the square
$$\begin{aligned} x^2 - 2(\mu+t\sigma^2)x + \mu^2 &= x^2 - 2(\mu+t\sigma^2)x + (\mu+t\sigma^2)^2 - (\mu+t\sigma^2)^2 + \mu^2 \\ &= (x-(\mu+t\sigma^2))^2 - 2t\sigma^2\mu - t^2\sigma^4 \end{aligned}$$
so
$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2(\mu+t\sigma^2)x+\mu^2}{2\sigma^2}}\, dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+t\sigma^2))^2-2t\sigma^2\mu-t^2\sigma^4}{2\sigma^2}}\, dx \\ &= e^{\mu t+\frac{\sigma^2t^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+t\sigma^2))^2}{2\sigma^2}}\, dx}_{=1 \text{ integral of PDF of } \mathrm{N}(\mu+t\sigma^2,\sigma^2)}. \end{aligned}$$

Negative Binomial: We first consider the case when $Y \sim \mathrm{NegBin}(1, p) = \mathrm{Geo}(p)$. In this case, using the geometric series (which exists for $|(1-p)e^t| < 1 \implies t < -\ln(1-p)$)
$$M_Y(t) = \sum_{x=0}^{\infty} e^{tx} p(1-p)^x = \sum_{x=0}^{\infty} p((1-p)e^t)^x = \frac{p}{1-(1-p)e^t}.$$
We recall that $X \sim \mathrm{NegBin}(k, p)$ is the sum of k independent $\mathrm{Geo}(p)$ random variables. Therefore, $X = Y_1 + \cdots + Y_k$ and the sum of independent random variables formula for MGFs give
$$M_X(t) = M_{Y_1+\cdots+Y_k}(t) = M_{Y_1}(t) \cdots M_{Y_k}(t) = M_Y^k(t) = \left(\frac{p}{1-(1-p)e^t}\right)^k,$$
which exists under the same condition as $M_Y(t)$.

Problem 1.31. Suppose that $M_X(t)$ is defined for all $t \in [-a, a]$ for some $a > 0$. Prove that

$$\mathbb{E}[X^k] = \frac{d^k}{dt^k} M_X(0) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0} \qquad \forall k \geq 0.$$

Solution 1.31. Assuming that we can interchange the differentiation and integration, we have

$$\frac{d^k}{dt^k} M_X(t) = \frac{d^k}{dt^k}\,\mathbb{E}[e^{tX}] = \mathbb{E}\left[\frac{d^k}{dt^k} e^{tX}\right] = \mathbb{E}[X^k e^{tX}]$$

Therefore,

$$\frac{d^k}{dt^k} M_X(0) = \mathbb{E}[X^k e^{0\cdot X}] = \mathbb{E}[X^k].$$

The interchange between the differentiation and expectation is justified by a result called the dominated convergence theorem. The technical assumption that $M_X(t)$ is defined for all $t \in [-a, a]$ for some $a > 0$ is required to find a dominating function. In the case when X has finite support, then the interchange of differentiation and expectation simply follows from the linearity of the differentiation operator.