1 Functions of Random Variables

We are often interested in (deterministic) functions of X. Let g be a function define let $Y := g(X)$. Notice that Y is a random variable which maps from the sample space $\mathcal{X}$ of X to the sample space $\mathcal{Y} = g(\mathcal{X})$ of Y, so it has a distribution given by

$$\mathbb{P}_Y(A) = \mathbb{P}(Y \in A) = \mathbb{P}(g(X) \in A) = \mathbb{P}(X \in g^{-1}(A)) = \mathbb{P}_X(g^{-1}(A))$$

given by the push-forward of $\mathbb{P}_X$ by g. This procedure allows us build complicated random variables from elementary ones, and can be extended to handle multiple random variables, which we will see in the coming weeks.

Remark 1. This is the exact same process we used to define $X : \Omega \to \mathbb{R}$ and its properties. We simply replace the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the sample space $(\mathcal{X}, \mathcal{B}, \mathbb{P}_X)$ and view $Y : \mathcal{X} \to \mathbb{R}$ to define the new sample space of $(\mathcal{Y}, \mathcal{B}, \mathbb{P}_Y)$ of Y.

1.1 Change of Variables Formula

If X is a continuous random variable, then there is a standard way to derive PDF of $Y = g(X)$ using calculus. Suppose that we know the PDF $f_X(x)$ of X. Our goal is to recover the the PDF $f_Y(y)$ of the random variable $Y = g(X)$. This can be done directly using the following steps

1. Use the support of X to find the support of $Y = g(X)$:

$$\mathcal{Y} = g(\mathcal{X}).$$

2. Compute the CDF of Y for $y \in \mathcal{Y}$ by expressing it in terms of the CDF of X:

$$F_Y(y) = \mathbb{P}(g(X) \le y) = \dots$$

   When the function g is not strictly increasing (or decreasing) over the support of X, then we must be careful when rewriting the inequality $\mathbb{P}(g(X) \le y)$.

3. Compute the PDF of Y by differentiating the CDF of Y,

$$f_Y(y) = F_Y'(y) \qquad y \in \mathcal{Y}.$$

When g is invertible, the above procedure gives us the change of density formula.

Theorem 1 (*Change of Variables Formula*)

Let X be a (absolutely) continuous random variable and g be invertible and differentiable with inverse g^{-1} on the support of $Y = g(X)$, then

$$f_Y(y) = |(g^{-1})'(y)| f_X(g^{-1}(y)) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)), \quad y \in \mathcal{Y}.$$

Remark 2. If X is discrete there is an analogous change of variables formula for the PMF of Y in terms of p_X and g^{-1},

$$p_Y(y) = \mathbb{P}(Y = y) = \sum_{x \in g^{-1}(Y)} \mathbb{P}(X = x) = \sum_{x \in g^{-1}(Y)} p_X(x), \quad y \in \mathcal{Y}$$

1.2 Example Problems

Problem 1.1. Let X be a continuous random variable with CDF $F_X(x) = \mathbb{P}(X \le x)$ and let $g : \mathbb{R} \mapsto \mathbb{R}$ be an increasing function with inverse g^{-1}. Compute $F_Y(y) = \mathbb{P}(Y \le y)$.

Solution 1.1. We can write the CDF of Y in terms of the CDF of X by

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

Notice that the CDF of $Y = g(X)$ is not $g(F_X(x))$.

Problem 1.2. Suppose $X \sim \text{Unif}(0,1)$, and that $Y = \frac{2}{X} - 1$. What is the support of Y?

Solution 1.2. Since $\text{supp}(X) = [0,1]$, we have

$$0 \le X \le 1 \implies \infty > \frac{2}{X} \ge 2 \implies \infty > \frac{2}{X} - 1 \ge 1$$

so $\text{supp}(Y) = [1, \infty)$.

Problem 1.3. Let X be a continuous random variable with the following pdf:

$$f_X(x) = \begin{cases} \frac{1}{4} & 0 < x \le 4, \\ 0 & \text{otherwise} \end{cases}$$

1. What is the range of Y
2. Find the CDF of X.
3. Let $Y = X^{-1}$ and find the PDF of Y.

Solution 1.3.

1. Outside the support of X, we have $F_X(x) = 0$ for $x < 0$ and $F_X(x) = 1$ for $x \ge 4$. For x in the support of X, i.e. $x \in (0,4]$
$$F_X(x) = \int_0^x \frac{1}{4} = \frac{x}{4}.$$
2. Notice that the range of X is $X(S) = (0,4)$ because $f_X(x) > 0$ for $x \in (0,4)$. Therefore, the range of $Y = X^{-1}$ is
$$0 < X \le 4 \implies \infty > X^{-1} \ge \frac{1}{4} \implies Y(S) = \Big[\frac{1}{4}, \infty\Big).$$
3. Outside the support of Y, we have $F_X(y) = 0$ for $y < \frac{1}{4}$. For y in the support of Y, i.e. $y \in \left[\frac{1}{4}, \infty\right)$
$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(X^{-1} \le y) = \mathbb{P}(X > \frac{1}{y}) = 1 - \mathbb{P}(X \le \frac{1}{y}) = 1 - F_X(y^{-1}) = 1 - \frac{1}{4y}.$$
To get the PDF of Y, we simply differentiate $F_Y(y)$ to conclude that
$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\Big(1 - \frac{1}{4y}\Big) = \frac{4}{y^2} \qquad \text{for } y \in \Big[\frac{1}{4}, \infty\Big).$$
and $f_Y(y) = 0$ for $y \notin \left[\frac{1}{4}, \infty\right)$.

Problem 1.4. Suppose a continuous random variable X has probability density function

$$f_X(x) = \begin{cases} 1 - |x| & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Find the CDF of X.
2. Let $Y = X^2$. Find the CDF Y.
3. Find the PDF of Y.

Solution 1.4.

1. We can rewrite the PDF as

$$f_X(x) = \begin{cases} 1 + x & -1 \le x < 0 \\ 1 - x & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

   Outside the support of X, we have $F_X(x) = 0$ for $x < -1$ and $F_X(x) = 1$ for $x > 1$. We have two cases for x in the support of X,

   (a) If $x \in [-1, 0]$

$$F_X(x) = \int_{-1}^{x} 1 + t \, dt = t + \frac{t^2}{2}\Big|_{-1}^{x} = x + \frac{x^2}{2} + \frac{1}{2}.$$

   (b) If $x \in [0, 1]$

$$F_X(x) = \int_{-1}^{0} 1 + t \, dt + \int_{0}^{x} 1 - t \, dt = \left(t + \frac{t^2}{2}\right)\Big|_{-1}^{0} + \left(t - \frac{t^2}{2}\right)\Big|_{0}^{x} = x - \frac{x^2}{2} + \frac{1}{2}.$$

   In summary,

$$F_X(x) = \begin{cases} 0 & x < -1 \\ x + \frac{1}{2}x^2 + \frac{1}{2} & -1 \le x < 0 \\ x - \frac{1}{2}x^2 + \frac{1}{2} & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

2. Notice that $\text{supp}(X) = [-1, 1]$. Therefore, the support of $Y = X^2$ is

$$-1 \le X \le 1 \implies 0 \le X^2 \le 1 \implies \text{supp}(Y) = [0, 1].$$

   Outside of the support, for $y < 0$ we have $F_Y(y) = 0$ and for $y > 1$ we have $F_Y(y) = 1$. In the support of Y, we have

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= \left(\sqrt{y} - \frac{1}{2}y + \frac{1}{2}\right) - \left(-\sqrt{y} + \frac{1}{2}y + \frac{1}{2}\right) \\ &= 2\sqrt{y} - y. \end{aligned}$$

   In summary,

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ 2\sqrt{y} - y & 0 \le y < 1 \\ 1 & 1 \le y \end{cases}$$

3. To get the PDF of Y, we simply differentiate $F_Y(y)$ to conclude that

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\Big(2\sqrt{y} - y\Big) = \frac{1}{\sqrt{y}} - 1 \qquad \text{for } y \in [0,1].$$

and $f_Y(y) = 0$ for $y \notin [0,1]$.

Remark 3. The function $g(x) = x^2$ is not increasing on the $\operatorname{supp}(X) = [-1,1]$ so we can't use the change of variables formula.

1.3 Proof of Key Results

Problem 1.5. Prove the change of variables formula in Theorem 1.

Solution 1.5.

Strictly Increasing: We first consider the case that g is strictly increasing. If g is strictly increasing, then

$$F_Y(y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X \le g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt.$$

Therefore, we can use the fundamental theorem of calculus and the chain rule to see that for points in the interior of the support of Y,

$$f_Y(y) = \frac{d}{dy}\int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) = \frac{1}{g'(g^{-1}(y))} f_X(g^{-1}(y)).$$

Since g is strictly increasing, $g' > 0$, so

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)), \quad y \in \operatorname{supp}(Y).$$

Strictly Decreasing: We now consider the case that g is strictly decreasing. If g is strictly decreasing, then

$$F_Y(y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X \ge g^{-1}(y)) = 1 - \mathbb{P}(X \le g^{-1}(y)) = 1 - 1 - \int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt.$$

Therefore, we can use the fundamental theorem of calculus and the chain rule to see that for points in the interior of the support of Y,

$$f_Y(y) = \frac{d}{dy}\left(1 - \int_{-\infty}^{g^{-1}(y)} f_X(t)\,dt\right) = -f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) = -\frac{1}{g'(g^{-1}(y))} f_X(g^{-1}(y)).$$

Since g is strictly increasing, $g' < 0$, so

$$f_Y(y) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)), \quad y \in \operatorname{supp}(Y).$$

2 Expected Value

The expectation of a random variable is the weighted average against its distribution. Statistically, the expected value is the best estimator that does not depend on the data (see problem 2.7). Mathematically, it is defined as the integral against the CDF.

Definition 1 (Expected Value)**.** If X is any random variable, then we can define

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x) \tag{1}$$

where F_X is the CDF of X. If $\mathbb{E}[|g(X)|] = \infty$ then we say that the expected value does not exist.

This is a *Riemann–Stieljes integral*, and it is defined in the natural way

$$\int_{-\infty}^{\infty} g(x) dF_X(x) = \lim_{\|\Pi\| \to 0} \sum_{i=1}^{n} g(x_i^*)[F_X(x_i) - F_X(x_i - 1)]$$

where the limit $\|\Pi\| \to 0$ means that the partition

$$-\infty < x_0 < \cdots < x_n < \infty$$

is taken so that $x_1 \to -\infty$, $x_n \to \infty$ and the increments $|x_i - x_{i-1}| \to 0$. The point x_i^* is an arbitrary sample point in the interval $[x_i - x_{i-1}]$.

If F_X is a step function, it gives us a summation at the jumps, and if F_X is absolutely continuous, it is integration against the density function.

Definition 2 (Expected Value — Discrete)**.** If X is a discrete random variable with PMF $p_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is a function,

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) p_X(x) = \sum_{x \in \mathcal{X}} g(x)\, \mathbb{P}(X = x)$$

provided it exists.

Definition 3 (Expected Value — Continuous)**.** If X is a continuous random variable with PDF $f_X(x)$, and $g : \mathbb{R} \to \mathbb{R}$ is a function, then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx,$$

provided it exists.

Remark 4. There exists an even more general notion of integration that ties everything together. We can alsto define the expected value as an integral on the underlying probability space Ω against the probability measure $\mathbb{P}$,

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega))\, d\,\mathbb{P}(\omega).$$

This is a *Lebesgue integral* and it is well defined if $\mathbb{E}[|g(X)|] < \infty$. It is the main focus of a course in measure theory. We can connect this integral to (1) through a change of variables.

2.0.1 Properties

1. Linearity: For any constants $a, b \in \mathbb{R}$,

$$\mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b$$

2. Linearity II:: Suppose that X and Y are random variables. Then

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

3. Jensen's Inequality: If g is convex then

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]).$$

4. Equivalent Formula: If $X \geq 0$, then

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq t)\, dt.$$

Remark 5. In general $\mathbb{E}[g(X)] \neq g[\mathbb{E}(X)]$. However, if g is a linear function or X is a constant random variable then equality holds.

Remark 6. In general $\mathbb{E}[XY] \neq \mathbb{E}[X]\,\mathbb{E}[Y]$. However, if X and Y are independent then equality holds.

2.1 Indicator Random Variables

The indicator functions provide a fundamental link between probability and expected values.

Definition 4 (Indicator Function)**.** Let $A \subset \Omega$ be an event. We say that $\mathbb{1}_A$ is the *indicator* random variable of the event A. $\mathbb{1}_A$ is defined by:

$$\mathbb{1}(\omega \in A) = \mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \in A^c \end{cases}.$$

Remark 7. The random variable $\mathbb{1}_A(\omega)$ is a Bernoulli random variable where a success is the occurrence of the event A.

2.2 Link Between Probabilities and Expected Values

The indicators link the concepts of expected values with the probability measure,

$$\mathbb{E}[1_A] = \mathbb{P}(A),$$

which follows from the simple fact that $\mathbb{1}_A \sim \text{Ber}(\mathbb{P}(A))$. This means that we can use indicator functions to write the theory of probability as the theory of integration, since the the probability of an event is precisely the integral of the indicator of the event against its probability distribution.

Naturally, the indicator functions behave quite similarly to probabilities and can be used as an alternative proof of the basic probability identities,

1. Complements: $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$ which implies that $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. Intersections: $\mathbb{1}_{A\cap B} = \mathbb{1}_A \mathbb{1}_B$, so if A and B are independent, then

$$\mathbb{P}(A\cap B) = \mathbb{E}[\mathbb{1}_{A\cap B}] = \mathbb{E}[1_A \mathbb{1}_B] = \mathbb{E}[\mathbb{1}_A]\,\mathbb{E}[\mathbb{1}_B] = \mathbb{P}(A)\,\mathbb{P}(B)$$

3. Inclusion – Exclusion: $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B}$ which implies that $\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cap B)$.
4. Union Bound: $\mathbb{1}_{A\cup B} \leq \mathbb{1}_A + \mathbb{1}_B$. which implies that $\mathbb{P}(A\cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$ by monotonicity of the expected value.

2.3 Example Problems

Problem 2.1. Consider a random variable X with PMF

$$p_X(1) = 0.3 \quad p_X(2) = 0.25 \quad p_X(3) = 0.2 \quad p_X(4) = 0.15 \quad p_X(5) = 0.1$$

What is $\mathbb{E}[X]$?

Solution 2.1. By definition,

$$\mathbb{E}[X] = 1 \cdot 0.3 + 2 \cdot 0.25 + 3 \cdot 0.2 + 4 \cdot 0.15 + 5 \cdot 0.1 = 2.5$$

Problem 2.2. Suppose X is a random variable satisfying $a \leq X(\omega) \leq b$ for all $\omega \in \Omega$. Show that $a \leq \mathbb{E}[X] \leq b$.

Solution 2.2. Since $a \leq X(\omega) \leq b$, we have $X(\Omega) \subseteq [a, b]$. There

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x p_X(x) \leq \sum_{x \in X(\Omega)} b p_X(x) = b \sum_{x \in X(\Omega)} p_X(x) = b$$

and

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x p_X(x) \geq \sum_{x \in X(\Omega)} a p_X(x) = a \sum_{x \in X(\Omega)} p_X(x) = a.$$

Problem 2.3. Suppose the discrete random variable X has PMF

$$p_X(-1) = 0.15, \quad p_X(0) = 2c, \quad p_X(1) = 0.5, \quad p_X(2) = 0.05, \quad p_X(3) = c.$$

where c is a constant making f a valid PMF.

1. What is $\mathbb{E}[X]$?
2. What is $\mathbb{E}[e^X]$?

Solution 2.3. Since the probabilities sum to 1, we must have

$$1 = \sum_{x=-1}^{3} p_X(i) = 3c + 0.7 \implies c = 0.1$$

1. By definition,

$$\mathbb{E}[X] = (-1) \cdot 0.15 + 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.05 + 3 \cdot 0.1 = 0.75$$

2. By the law of the unconscious statistician

$$\mathbb{E}[e^X] = e^{-1} \cdot 0.15 + e^0 \cdot 0.2 + e^1 \cdot 0.5 + e^2 \cdot 0.05 + e^3 \cdot 0.1 \approx 3.992$$

Problem 2.4. Consider a random variable X with PMF $p_X(x) = \frac{1}{x}$ for $x = 2, 4, 8, 16, \dots$ and 0 otherwise.

1. Show that $\sum_{\text{all x}} p_X(x) = 1$.

2. What is $\mathbb{E}[X]$?

Solution 2.4.

1. The range of X is 2^n where $n \geq 1$. We have by the geometric series that

$$\sum_{\text{all x}} p_X(x) = \sum_{n\geq 1} \frac{1}{2^n} = \sum_{n \geq 0} \frac{1}{2^n} - 1 = \frac{1}{1-\frac{1}{2}} - 1 = 1.$$

2. For the expected value, we have

$$\mathbb{E}[X] = \sum_{n \geq 1} 2^n p_X(2^n) = \sum_{n=1}^{\infty} 2^n \left(\frac{1}{2^n}\right) = \sum_{n=1}^{\infty} 1 = \infty,$$

so the expected value is infinite, and does not exist in the traditional sense.

Problem 2.5. Suppose the waiting time X until Mukhtar's next bus arrives follows an exponential distribution with parameter $\theta = 1$. What's the waiting time w so that Mukhtar doesn't have to wait longer than w with probability 50%?

Solution 2.5. We know

$$F_X(x) = \mathbb{P}(X \leq x) = \int_0^x e^{-t}\, dt = 1 - e^{-x}$$

for $x \geq 0$ and 0 otherwise. This function is strictly increasing on $x > 0$, so we can use the classical inverse. We want w such that $\mathbb{P}(X \leq w) = 0.5$, or

$$F_X(w) = 0.5 \Leftrightarrow 1 - e^{-w} = 0.5 \Leftrightarrow w = \log(2) \approx 0.693.$$

So with probability 50% Mukhtar won't have to wait longer than $\log(2)$. In other words, $\log(2)$ is the median of the distribution of X.

Remark 8. We can repeat this computation for general $\theta \neq 1$. The median in this case will be given by $\theta \log(2)$. Since $\log(2) < 1$ this implies that the median of the exponential is always below the mean.

Problem 2.6. Uranium 238 emits particles measured by a Geiger counter at a rate of 50 per second. Assume that the number of particles measured by a Geiger counter follows a Poisson process. Let X denote the amount of time in seconds between when the first and second particles are measured. Find $\mathbb{E}[X]$

Solution 2.6. Intuitively, since the rates are homogeneous and the increments are independent, the time between the first and second particle is equal in distribution to the time between the first particle. Therefore, the waiting time parameter θ is $\frac{1}{50}$ seconds per occurrence, so $X \sim \text{Exp}(50^{-1})$. Therefore, $\mathbb{E}[X] = \theta = \frac{1}{50}$.

We carefully explain how the homogeneous and independence property are used in this problem. Let X be the time of the first particle, Y be the time from the first to the second particle, and $N(t)$ be the number of particles by time t. Clearly, $Z = X + Y$ is the time of the second particle.

We know $X \sim \text{Exp}(50^{-1})$ and $N(t) \sim \text{Poi}(50t)$. Using the same logic as in the derivation of the Poisson process (see Problem (**??**)),

$$F_Z(t) = \mathbb{P}(X+Y \le t) = 1 - \mathbb{P}(X+Y > t) = 1 - P(N(t) \le 1) = 1 - (e^{-50t} + e^{-50t}(50t))$$

So $Z = X + Y$ has density

$$f_Z(t) = \frac{d}{dt}\mathbb{P}(X+Y \le t) = -\frac{d}{dt}\left[e^{-50t}(1+50t)\right] = t50^2 e^{-50t}$$

Therefore, by an integration by parts

$$\mathbb{E}[Z] = \int_{-\infty}^{\infty} z \cdot f_Z(z) dz = 50 \int_0^{\infty} z^2 \cdot 50 e^{-50z} dz = 50 \frac{2}{50^2} = \frac{2}{50}$$

and the linearity of expectation implies that

$$\mathbb{E}[Y] = \mathbb{E}[Z - X] = \mathbb{E}[Z] - \mathbb{E}[X] = \frac{2}{50} - \frac{1}{50} = \frac{1}{50}.$$

2.4 Proofs of Key Results

Problem 2.7. For any constant c, show that

$$\mathbb{E}[(X-c)^2] \ge \mathbb{E}[(X - \mathbb{E}[X])^2].$$

In particular, the expected value is the constant that minimizes the mean squared error. Furthermore, the value of the mean squared error is given by the variance.

Solution 2.7. This proof follows directly from the properties of the expected value. We have

$$\mathbb{E}[(X-c)^2] = \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - c)^2] = \mathbb{E}[X^2] - 2c\,\mathbb{E}[X] + c^2.$$

Therefore,

$$\frac{d}{dc}\mathbb{E}[(X-c)^2] = -2\,\mathbb{E}[X] + 2c.$$

The critical point of this function is attained when $c = \mathbb{E}[X]$, and this point is a global minimum since $c \mapsto \mathbb{E}[(X-c)^2]$ is an upward facing parabola.

Alternative Proof: We can also do this proof without calculus. This proof follows directly from the properties of the expected value. By adding and subtracting $\mathbb{E}[X]$, we see that

$$\begin{aligned}\mathbb{E}[(X-c)^2] &= \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - c)^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X])^2] + \mathbb{E}[(\mathbb{E}[X] - c)^2] + 2\,\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[X] - c)]\end{aligned}$$

Since $\mathbb{E}[X] - c$ is not random, we see that the cross terms vanish

$$\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[X] - c)] = (\mathbb{E}[X] - c)\,\mathbb{E}[(X - \mathbb{E}[X])] = (\mathbb{E}[X] - c)(\mathbb{E}[X] - \mathbb{E}[X]) = 0.$$

Since $\mathbb{E}[(\mathbb{E}[X] - c)^2] \ge 0$, we conclude that

$$\mathbb{E}[(X-c)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2] + \mathbb{E}[(\mathbb{E}[X] - c)^2] \ge \mathbb{E}[(X - \mathbb{E}[X])^2]$$

as required.

Problem 2.8. Prove the law of the unconscious statistician. That is, if we define $Y = g(X)$, then

$$\mathbb{E}[Y] = \sum_{x \in Y(\Omega)} y p_Y(y)$$

is the same as

$$\mathbb{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) p_X(x).$$

Solution 2.8. This follows from a change of variables and a rearrangement of a sum. Given a function g, let

$$D_y = g^{-1}(\{y\}) = \{x : g(x) = y\}.$$

If we let $Y = g(X)$ then

$$p_Y(y) = \mathbb{P}(g(X) = y) = \sum_{x \in D_y} p_X(x).$$

and $Y(\Omega) = g(X(\Omega))$ since

$$Y(\Omega) = \{Y(\omega) : \omega \in \Omega\} = \{g(X(\omega)) : \omega \in \Omega\}.$$

Therefore,

$$\mathbb{E}[Y] = \sum_{y \in Y(\Omega)} y f_Y(y) = \sum_{y \in Y(\Omega)} y \sum_{x \in D_y} p_X(x) = \sum_{y \in Y(\Omega)} \sum_{x \in D_y} g(x) p_X(x) = \sum_{x \in X(\Omega)} g(x) p_X(x).$$

Since $(D_y)_{y \in Y(\Omega)}$ forms a partition of $X(\Omega)$.

Problem 2.9. Prove the linearity property of expectation

Solution 2.9. This follows form the linearity of summation. By the law of the unconscious statistician with $g(x) = ax + b$, we have

$$\mathbb{E}[aX + b] = \sum_{x \in X(\Omega)} (ax + b) p_X(x) = a \sum_{x \in X(\Omega)} x p_X(x) + b \sum_{x \in X(\Omega)} p_X(x) = a\,\mathbb{E}[X] + b$$

since $\sum_{x \in X(\Omega)} p_X(x) = 1$.

Problem 2.10. Prove the equivalent formula for the expected value when X is non-negative.

Solution 2.10. If $x \geq 0$, we can write it as

$$x = \int_0^x dt = \int_0^\infty \mathbb{1}(t \leq x)\, dt.$$

Using this fact in the definition of the expected value gives us

$$\begin{aligned}
\mathbb{E}[X] = \sum_{x \in X(\Omega)} x\, \mathbb{P}(X = x) &= \sum_{x \in X(\Omega)} \int_0^\infty \mathbb{1}(t \leq x)\, \mathbb{P}(X = x)\, dt \\
&= \int_0^\infty \sum_{x \in X(\Omega)} \mathbb{1}(t \leq x)\, \mathbb{P}(X = x)\, dt \\
&= \int_0^\infty \sum_{x \in X(\Omega): x \geq t} \mathbb{P}(X = x)\, dt \\
&= \int_0^\infty \mathbb{P}(X \geq t)\, dt.
\end{aligned}$$

The interchange of the sum and integral can always be done if $X(\Omega)$ is finite. If $|X(\Omega)|$ is infinite, then it is also true and it can be justified by the monotone convergence theorem.

Problem 2.11. If $X \sim \text{DUnif}[a,b]$ then $\mathbb{E}[X] = \frac{a+b}{2}$.

Solution 2.11. If $X \sim \text{DUnif}[a,b]$ then the sum of positive integers implies

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

Alternative Solution: If $X \sim \text{DUnif}[0,n]$ 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 $X \sim \text{DUnif}[a,b] \sim a + \text{DUnif}[0,b-a]$. Therefore, if we let $Y \sim \text{DUnif}[0,b-a]$, 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 2.12. If $X \sim \text{Hyp}(N,r,n)$, then $\mathbb{E}[X] = r\frac{n}{N}$.

Solution 2.12. 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 $X \sim \text{Hyp}(N,r,n)$ denotes the number of blue balls we drew from a sample of n balls without replacement.

We label the blue balls $1,\dots,r$ 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 2.13. If $X \sim \text{Bin}(n,p)$ then $\mathbb{E}[X] = np$.

Solution 2.13. 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 $X \sim \text{Bern}(p)$ 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 $X \sim \text{Bern}(p)$ (the number of successes in n trials is equal to the sum of n successful trials), linearity implies that

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

Problem 2.14. If $X \sim \text{NegBin}(k,p)$, show that $\mathbb{E}[X] = \frac{k(1-p)}{p}$.

Solution 2.14. 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 2.15. If $X \sim \text{Poi}(\lambda)$, show that $\mathbb{E}[X] = \lambda$.

Solution 2.15. 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 2.16. If $X \sim \text{Unif}(a, b)$ show that $\mathbb{E}[X] = \frac{a+b}{2}$.

Solution 2.16. We can directly compute

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

Problem 2.17. Let $n \ge 1$. Show that

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

Solution 2.17. This follows from repeated integration by parts. Let $n \ge 1$, 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

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

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

Remark 9. 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 2.18. Prove that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Solution 2.18. We have

$$\Gamma\Big(\frac{1}{2}\Big) = \int_0^\infty y^{\frac{1}{2}-1} e^{-y} dy = \int_0^\infty y^{-\frac{1}{2}} e^{-y} dy$$

We can do the change of variables $y = \frac{x^2}{2}$, $dy = x\,dx$ so

$$\Gamma\Big(\frac{1}{2}\Big) = \sqrt{2} \int_0^\infty e^{-\frac{x^2}{2}}\,dx.$$

By Fubini's theorem and a change of variables into polar coordinates

$$\begin{aligned}\Gamma\Big(\frac{1}{2}\Big)^2 = 2\bigg(\int_0^\infty e^{-\frac{x^2}{2}}\,dx\bigg)^2 = 2\int_0^\infty e^{-\frac{x^2}{2}}\,dx \int_0^\infty e^{-\frac{y^2}{2}}\,dy &= 2\int_0^\infty \int_0^\infty e^{-\frac{x^2+y^2}{2}}\,dxdy \\ &= 2\int_0^{\frac{\pi}{2}} e^{-\frac{r^2}{2}} r\,drd\theta \\ &= 2\int_0^{\frac{\pi}{2}} d\theta = \pi.\end{aligned}$$

So $\Gamma\Big(\frac{1}{2}\Big) = \sqrt{\pi}$.

Remark 10. A modification of this argument can be used to compute the normalization constant of a standard Gaussian random variable.

Problem 2.19. If $X \sim \mathrm{Exp}(\theta)$ show that $\mathbb{E}[X] = \theta$.

Solution 2.19. We use the change of variable $x = y\theta$ with $dx = \theta dy$

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

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

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

Problem 2.20. If $X \sim \mathrm{N}(\mu, \sigma^2)$ show that

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

Solution 2.20. 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.

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

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

Problem 2.21. Show the following properties of an indicator function

1. Complements: $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$
2. Intersections: $\mathbb{1}_{A \cap B} = \mathbb{1}_A \mathbb{1}_B$
3. Inclusion – Exclusion: $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B}$
4. Union Bound: $\mathbb{1}_{A \cup B} \leq \mathbb{1}_A + \mathbb{1}_B$.

Solution 2.21. The proofs are quite straightforward and essentially follow from the facts that $1-0 = 1$ and $1 \cdot 1 = 1$.

1. Complements: On one side we have

$$\mathbb{1}_{A^c} = \begin{cases} 1 & x \in A^C \\ 0 & x \in A \end{cases}.$$

On the other hand, we have

$$1 - \mathbb{1}_A = \begin{cases} 1-1 & x \in A \\ 1-0 & x \in A^c \end{cases} = \begin{cases} 1 & x \in A^C \\ 0 & x \in A \end{cases},$$

so both sides are equivalent.

2. Intersections: On one side, we have

$$\mathbb{1}_{A \cap B} = \begin{cases} 1 & x \in A \text{ and } x \in B \\ 0 & \text{otherwise} \end{cases}.$$

On the other hand, we have

$$\mathbb{1}_{A \cap B} = \begin{cases} 1 \cdot 1 & x \in A \text{ and } x \in B \\ 0 & \text{otherwise} \end{cases}$$

so both sides are equivalent.

The rest of the identities are verified similarly.