Probability Lemmas, Definitions, Axioms and Theorems

Contents

1 Lemmas

1.1 Arrangements of n objects

The number of arrangements of the n objects

$$\underbrace{x_1, x_1, x_1 \ldots, x_1}_{m_1 \text{ times}}, \underbrace{x_2, x_2, x_2 \ldots, x_2}_{m_2 \text{ times}}, \ldots, \underbrace{x_k, x_k, x_k \ldots, x_k}_{m_k \text{ times}}$$

where x_i appears m_i times is

$$\frac{n!}{m_1!m_2!\ldots m_k!} \tag{1.1.1}$$

If just 2 tpes of object are present, then the binomial coefficient may be used. $m_1+m_2 = n \Rightarrow (1.1.1) = \binom{n}{m_1} = \binom{n}{m_2}$

1.2 Vandermonde's Identity

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} \tag{1.2.1}$$

where $\binom{m}{j} = 0$ for $j > m$.

Proof. For example, suppose we choose a committee consisting of k people from a group of m men and n women. There are $m + n$ ways of doing this which is the left-hand side of (1.2.1).

Now the number of men in the committee is some $j \in \{0, 1, \ldots, k\}$ so it contains $k - j$ women. The number of ways of choosing the j men is $\binom{m}{j}$ and for each such choice there are $\binom{n}{k-j}$ choices for the women who make up the rest of the committee. So there are $\binom{m}{j}\binom{n}{k-j}$ committees with exactly j men and summing over j we get that the total number of committees is given by the right-hand side of (1.2.1) □

1.3 Probability Space

A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$where

- Ω is the sample space
- $\mathcal{F}$ is a collection of subsets of Ω, called events, satisfying axioms F_1, F_2, F_3 below,
- $\mathbb{P}$ is a probability measure which is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying axioms P_1 to P_4 below.

1.4 Conditional Probability

If $\mathbb{P}(B) > 0$ the conditional probability of A given B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

1.5 Conditional Distribution

Suppose B is an event such that $\mathbb{P}(B) > 0$. Then the conditional distribution of $X \mid B$ is $\mathbb{P}(X = x \mid B)$ for $x \in \mathbb{R}$.

$$p_{X|B}(x) = \frac{\mathbb{P}(\{X = x\} \cap B)}{\mathbb{P}(B)}$$

The conditional expectation of $X \mid B$ is

$$\mathbb{E}[X \mid B] = \sum_{x \in \operatorname{Im} X} x\mathbb{P}(X = x \mid B)$$

1.6 Conditional Expectation

The conditional expectation of Y given $X = x$ is $\mathbb{E}[Y \mid X = x] = \sum_y y p_{Y|X=x}(y)$ whenever this exists.

1.7 Independence

For a pair of events A and B, A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

The family of events $\{A_i, i \in I\}$ where I is an index set; is independent if

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \text{ for all finite subsets } J \in I$$

A family $\{A_i, i \in I\}$ is pairwise independent if $\mathbb{P}(A_i, i \in I) = \mathbb{P}(A_i)\mathbb{P}(A_j), \forall i \neq j$. Pairwise independence does not imply independence.

Note: for the independent events A and B, knowing $\mathbb{P}(A)$ does not imply knowledge of $\mathbb{P}(B)$

For a pair of discrete random variables X and Y on the same probability space, they are inde3pendent if $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \forall x, y \in \mathbb{R}$. i.e., X and Y are independent iff $\{X = x\}$ and $\{Y = y\}$ are independent events *for all* choices of $x, y \in \mathbb{R}$

1.8 Discrete Random Variables

A discrete random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is simply a function $X : \Omega \to \mathbb{R}$ with the properties that:

- $\operatorname{Im}(X) = \{X(\omega) : \omega \in \Omega\}$ is a finite or countably infinite subset of $\mathbb{R}$. i.e. $\operatorname{Im}(X)$ can be written as $\{x_0, x_1, x_2 \ldots\}$. Usually $\operatorname{Im}(X)$ will be $\mathbb{N}$ or a subset of $\mathbb{N}$.
- $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$. This says that if $\{\omega \in \Omega : X(\omega) = x\}$ is an event then we can assign a probability to it. $\{\omega \in \Omega : X(\omega) = x\}$ is usually abbreviated to $\{X = x\}$. $\mathbb{P}(X = x)$ is used as a shorthand for $\mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$
- Let $f : \mathbb{R} \to \mathbb{R}$, then if X is a discrete random variable, then $Y := f(X)$ is also a discrete random variable. If $\operatorname{Im}(X) = \{x_1, x_2, \ldots\}$ then $\operatorname{Im}(f(x)) = \{f(x_1), f(x_2), \ldots\}$.

1.9 Probability Mass Function

The probability mass function of a discrete random variable X is the function $p_X : \mathbb{R} \to [0,1]$ defined by $p_X(x) = \mathbb{P}(X = x)$. If $x \notin \mathrm{Im}(X)$ then $p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\emptyset) = 0$. Also if $\mathrm{Im}(X) = \{x_0, x_1, \ldots\}$ then

$$\sum_{i \in \mathrm{Im}(X)} p_X(x) = \sum_{i \geq 0} p_X(x_i)$$
$$= \sum_{i \geq 0} \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i\})$$

events are disjoint

$$= \mathbb{P}\left(\bigcup_{i \geq 0} \{\omega \in \Omega : X(\omega) = x_i\}\right)$$

every event in Ω mapped to $x_0, x_1, \ldots$

$$= \mathbb{P}(\Omega)$$

By (P_1)

$$= 1$$

1.10 Classical Distributions

1.10.1 Bernoulli Distribution

X has the Bernoulli Distribution with parameter p, where $p \in [0,1]$ if $\mathbb{P}(X = 0) = 1 - p$ and $\mathbb{P}(X = 1) = p$. As $1-p+p = 1$, X can only take the values 0 and 1. $p_X(0) = 1-p$ and $p_X(1) = p$. $\mathbb{1}_A$ (the indicator function of A) is an example of a Bernoulli random variable with $p = \mathbb{P}(A)$, constructed on an explicit probability space. We write $X \sim \mathrm{Ber}(p)$ to indicate a discrete random variable X modelled by $\mathrm{Ber}(p)$. Can be used to model flipping a (possibly unfair) coin.

1.10.2 Binomial Distribution

X has a binomial distribution with parameters n and p, where $n \in \mathbb{Z}, p \in [0,1]$ if $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for some constant k where $0 \leq k \leq n$. We write $X \sim \mathrm{Bin}(n, p)$. For example, the number of heads from n independent trials where $\mathbb{P}(\{\text{heads}\}) = p$

1.10.3 Geometric Distribution

X has a geometric distribution with parameter $p \in [0,1]$ if $\mathbb{P}(X = k) = (1-p)^{k-1} p, k \in \mathbb{Z}$. We write $X \sim \mathrm{Geom}(p)$. It is used to model the number of independent trials needed until the first success occurs, where $\mathbb{P}(\{\text{success}\}) = p$

Alternative Geometric Distribution

Is used to model the number of failures before first success. $\mathbb{P}(Y = k) = (1-p)^k p, k \geq 0$

1.10.4 Poisson Distribution

X has a Poisson distribution with parameter $\lambda \geq 0, \lambda \in \mathbb{R}$ if $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{Z}_0$. We write this as $X \sim \mathrm{Po}(\lambda)$ and is used to model events occuring at a constant average rate λ.

1.11 Expectation

The expectation (or mean) of a variable X is

$$\mathbb{E}[X] = \sum_{x \in \mathrm{Im}\, X} x \mathbb{P}(X = x)$$

provided that $\sum_{x \in \mathrm{Im}\, X} |x| \mathbb{P}(X = x) < \infty$. If $\mathbb{E}[X]$ diverges, then it does not exist.

1.12 Moments

The k^{th} moment of a variable X is $m_k(X) = \mathbb{E}\left[X^k\right]$ given that it exists.

1.13 Joint Distributions

Suppose we consider the discrete random variables X and Y on the same probability space. In order to under stand the distribution of this pair, we must specify $\mathbb{P}(\{X = x\} \cap \{Y = y\})$ *for all* $x, y \in \mathbb{R}$. It is not enough to specify $\mathbb{P}(X = x)$ and $\mathbb{P}(Y = y)$ as X and Y may not be independent.

1.14 Joint Probability Mass Function

The joint probability mass function of X and Y is given by $P_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$. $P_{X,Y}(x, y) > 0, \forall x, y \in \mathbb{R}$ and $\sum_x \sum_y p_{X,Y}(x, y) = 1$.

1.15 Marginal Distribution

The marginal distribution of X is $p_X(x) = \mathbb{P}(X = x) = \sum_y p_{X,Y}(x, y)$. Similarly for Y, the marginal distribution is given as $\sum_x p_{X,Y}(x, y)$.

1.16 Covariance

Given X and Y *not* independent, their covariance is $\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. It is therefore easy to see that $\mathrm{cov}(X, X) = \mathrm{var}(X)$. If two variables are independent (hence $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$) then $\mathrm{cov}(X, Y) = 0$. However, the reverse is not true.

1.17 Multivariate Joint Distributions

By extending the principal of joints distributions, $p_{X_1,X_2,\ldots,X_k}(x_1, x_2, \ldots) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$ for $x_1, x_2, \ldots, x_n \in \mathbb{R}$

1.18 Multivariate Independence

A family $\{X_i : i \in I\}$ of random variables is independent for all finite subsets of $J \subseteq I$ and all collections $\{x_i : i \in J\}$ if

$$\mathbb{P}\left(\bigcap_{i \in J}\{X_i = x_i\}\right) = \prod_{i \in J} \mathbb{P}(X_i = x_i)$$

1.19 Variance

The variance of a variable is the measure of spread around its mean. For a discrete random variable X, its variance is defined by: $\mathrm{var}(X) = \mathbb{E}\left[X - \mathbb{E}\left[X^2\right]\right]$ when $\mathbb{E}[X]$ exists. This can also be re-written as $\mathbb{E}[f(X)]$ for $f(X) = (X - \mathbb{E}[X])^2$, $\mathbb{E}[X] \in \mathbb{R}, f : \mathbb{R} \to \mathbb{R}$. Since $f(X) \geq 0$, $\mathrm{var}(X) \geq 0$. By letting $\mu = \mathbb{E}[X]$ and substitute in $f(X) = (X - \mu)^2$. The standard deviation is often preferred as a measure of spread as it preserves linearity when scaling.

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}\left[X^2 - 2\mu X + \mu^2\right] \\
&= \sum_{x \in \mathrm{Im}\, X} \left(x^2 - 2\mu x + \mu^2\right) p_X(x) \\
&= \sum_{x \in \mathrm{Im}\, X} x^2 p_X(x) - 2\mu \sum_{x \in \mathrm{Im}\, X} x p_X(x) + \mu^2 \sum_{x \in \mathrm{Im}\, X} p_X(x) \\
&= \mathbb{E}\left[X^2\right] - 2\mu \mathbb{E}[X] + \mu^2 \\
\overset{\mu = \mathbb{E}[X]}{\Rightarrow} &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2
\end{aligned}$$

2 Axioms

- $$\emptyset \in \mathcal{F}, \Omega \in \mathcal{F} \tag{F$_1$}$$
  The empty set and the sample space are both in the list of possible outcomes. i.e. nothing happens, or something happens.
- $$\text{If } A, B \in \mathcal{F}, \text{ then } A^c \in \mathcal{F} \text{ and } A \cup B \in \mathcal{F} \tag{F$_2$}$$
  For any set in the set of possible outcomes, its complement is in the list of outcomes. For any two sets in the set of possible outcomes, their unison is in the list of outcomes.
- $$\text{If } A_i \in \mathcal{F} \text{ for } i \geq 1, \text{ then } \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \tag{F$_3$}$$
  For any list of sets in the set of possible outcomes, the union of the list of sets is a possible outcome.
- $$\forall A \in \mathcal{F}, \mathbb{P}(A) \geq 0 \tag{P$_1$}$$
  Any outcome in the list of possible outcomes has a chance of occuring (probability of greater than 0) or no chance of occuring (probability of 0)
- $$\mathbb{P}(\Omega) = 1 \tag{P$_2$}$$
  The probability of any outcome occurring is 1. (Something will always happen)
- $$\text{If } A, B \in \mathcal{F} \text{ and } A \cap B = \emptyset \text{ then } \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \tag{P$_3$}$$
  If two sets of outcomes are disjoint (mutually exclusive), then the probability of either is the sum of their individual probabilities.

- $$\text{If } A_i \in \mathcal{F} \text{ for } i \geq 1 \text{ and } A_i \cap A_j = \emptyset \text{ for } i \neq j \text{ then } \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \tag{P$_4$}$$

  For a set of mutually exlusive events, the probability of any one happening is the same as the sum of all the individual events.

3 Theorems

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

- Let X be a discrete random variable such that $\mathbb{E}[X]$ exists.
  - If X is non-negative, then so is $\mathbb{E}[X]$
  - For $a, b \in \mathbb{R}$, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
  - *Proof.* $\operatorname{Im} X \subseteq [0, \infty)$ so $\sum_{x \in \operatorname{Im} X} x\mathbb{P}(X = x)$ is a sum whose terms are all non-negative, which must give a non-negative solution. □
  - *Proof.* Let $f(X) = aX + b$. Then

$$\begin{aligned}
\sum_{x \in \operatorname{Im} X} (ax + b)\mathbb{P}(f(X) = ax + b) &= \sum_{x \in \operatorname{Im} X} ax\mathbb{P}(f(X) = ax + b) \\
&\quad + \sum_{x \in \operatorname{Im} X} b\mathbb{P}(f(X) = ax + b) \\
&= a \sum_{x \in \operatorname{Im} X} x\mathbb{P}(f(X) = aX + b) + b \\
&= a\mathbb{E}[X] + b
\end{aligned}$$

□

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

  Proof.

$$\begin{aligned}
A \cup A^c = \Omega &\text{ and } A \cap A^c = \emptyset \\
\text{By (P}_3) &\Rightarrow \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c) \\
\text{By (P}_2) &\Rightarrow 1 = \mathbb{P}(A) + \mathbb{P}(A^c) \\
&\Rightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A)
\end{aligned}$$

□

- If $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

  Proof. Since $A \subseteq B, B = A \cup (B \cap A^c)$. Since $B \cap A^c \subseteq A^c$, it must be disjoint from A. So by (P$_3$), $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$. Since by (P$_1$), $\mathbb{P}(B \cap A^c) \geq 0$, we thus have $\mathbb{P}(B) \geq \mathbb{P}(A)$. □

- Suppose $\mathbb{P}(B) > 0$. Define the new function $\mathbb{Q} : \mathcal{F} \to \mathbb{R}$ then $\mathbb{Q}(A) = \mathbb{P}(A|B)$. Then $(\Omega, \mathcal{F}, \mathbb{Q})$ is also a probability space.

  Proof. We only need to check that $\mathbb{Q}$ satisfies (P$_1$) to (P$_4$) as we are using the same $\mathcal{F}$. (P$_1$)

$$\mathbb{Q}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0$$

  (P$_2$)

$$\mathbb{Q}(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)}$$

  (P$_3$) and (P$_4$) have the same proof, so we just do (P$_4$): For disjoint events $A_1, A_2, \ldots,$

$$\begin{aligned}
\mathbb{Q}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \frac{\mathbb{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathbb{P}(B)}
\end{aligned}$$

  (because $A_i \cap B, i \geq 1$, are disjoint)

$$\begin{aligned}
&= \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{\infty} \mathbb{Q}(A_i)
\end{aligned}$$

□

- **Law of Total Probability or Partition Theorem** If $\{B_1, B_2 \ldots\}$ if a partitions of Ω such that $\mathbb{P}(B_i) > 0, \forall i \geq 1$ then $\mathbb{P}(A) = \sum_{i \geq 1} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$

  Proof.

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega)$$

Since $\bigcup_{i\geq 1} B_i = \Omega$

$$= \mathbb{P}\left(A \cap \bigcup_{i\geq 1} B_i\right)$$
$$= \mathbb{P}\left(\bigcup_{i\geq 1} (B_i \cap A)\right)$$

As $A \cap B_i = \emptyset$ for $i \geq 1$, then by (P_4)

$$= \sum_{i\geq 1} \mathbb{P}(A \cap B_i)$$

Through application of the multiplication rule

$$= \sum_{i\geq 1} \mathbb{P}(A \cap B_i)\,\mathbb{P}(B_i)$$

$\square$

- If A and B are independent, then so are the events A and B^c, as well as A^c and B^c.

  Proof. $A = (A\cap B)\cup(A\cap B^c)$ and $(A\cap B)\cap(A\cap B^c) = \emptyset$. So, by ($\text{P}_3$):

$$\begin{aligned}\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B)\\ &= \mathbb{P}(A)(1 - \mathbb{P}(B))\\ &= \mathbb{P}(A)\,\mathbb{P}(B^c)\end{aligned}$$

$\square$

  Proof. Apply above to the events B^c and A $\square$

- **Bayes' Theorem** Suppose that $\{B_1, B_2, \ldots\}$ is a partition of Ω and $\mathbb{P}(B_i) > 0, \forall i \geq 1$. Also that $\mathbb{P}(A) > 0$. Then

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\,\mathbb{P}(B_k)}{\sum_{i\geq 1} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}$$

  Proof.

$$\begin{aligned}\mathbb{P}(B_k \mid A) &= \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)}\\ &= \frac{\mathbb{P}(A \mid B_k)\,\mathbb{P}(B_k)}{\mathbb{P}(A)}\end{aligned}$$

  By applying the partition theorem for $\mathbb{P}(A)$

$$= \frac{\mathbb{P}(A \mid B_k)\,\mathbb{P}(B_k)}{\sum_{i\geq 1} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}$$

$\square$

- If $f : \mathbb{R} \to \mathbb{R}$ then $\mathbb{E}\left[f(X) = \sum_{x\in \operatorname{Im} X} f(x)\mathbb{P}(X = x)\right]$ if $\sum_{x\in \operatorname{Im} X} |f(x)|\,\mathbb{P}(X = x) < \infty$

  Proof. Let $A = \{y : y = f(x) \text{ for some } x \in \operatorname{Im} X\}$. Then

$$\begin{aligned}\sum_{x\in \operatorname{Im} X} f(x)\mathbb{P}(X = x) &= \sum_{y\in A}\left(\sum_{\substack{x\in \operatorname{Im} X\\ y=f(x)}} f(x)\mathbb{P}(X = x)\right)\\ &= \sum_{y\in A} y \sum_{\substack{x\in \operatorname{Im} X\\ y=f(x)}} \mathbb{P}(X = x)\\ &= \sum_{y\in A} y\mathbb{P}(f(X) = y)\\ &= \mathbb{E}[f(X)]\end{aligned}$$

$\square$

- For $a, b \in \mathbb{R}$, $\operatorname{var}(aX + b) = a^2 \operatorname{var}(X)$

  Proof.

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2\\ \operatorname{var}(aX + b) &= \mathbb{E}\left[(aX + b)^2\right] - \mathbb{E}[aX + b]^2\\ &= \mathbb{E}\left[a^2X^2 + 2bX + b^2\right] - a^2\mathbb{E}[X]^2 - 2\mathbb{E}[X] + b^2\\ &= \sum_x \left(a^2x^2 + 2bx + b^2\right) - a^2\mathbb{E}[X]^2 - 2\mathbb{E}[X] + b^2\\ &= \sum_x \left(a^2x^2\right) + 2b\mathbb{E}[X] - 2b\mathbb{E}[X] + b^2 - b^2\\ &= a^2\mathbb{E}\left[X^2\right] - a^2\mathbb{E}[X]^2\\ &= a^2 \operatorname{var}(X)\end{aligned}$$

$\square$

- **Law of Total Probability for Expectations** If $\{B_1, B_2, \ldots\}$ is a partition of Ω such that $\mathbb{P}(B_i) > 0\forall i$, then $\mathbb{E}[X] = \sum_{i\geq 1} \mathbb{E}[X \mid B_i]\,\mathbb{P}(B_i)$ whenever $\mathbb{E}[X]$ exists.

  Proof.

$$\begin{aligned}\mathbb{E}[X] &= \sum_x x\mathbb{P}(X = x)\\ &= \sum_x x \sum_{i\geq 1} \mathbb{P}(X = x \mid B_i)\,\mathbb{P}(B_i)\\ &= \sum_x \sum_{i\geq 1} x\mathbb{P}(X = x \mid B_i)\,\mathbb{P}(B_i)\\ &= \sum_{i\geq 1} \mathbb{P}(B_i) \sum_x x\mathbb{P}(X = x \mid B_i)\\ &= \sum_{i\geq 1} \mathbb{P}(B_i)\,\mathbb{E}[X \mid B_i]\end{aligned}$$

$\square$

- **Linearity of $\mathbb{E}$** Given the constants $a, b \in \mathbb{R}$, then $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ whenever $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ exist.

Proof. Let $f(X,Y) = aX + bY$

$$
\begin{aligned}
&\Rightarrow \sum_x \sum_y (ax+by) p_{X,Y}(x,y) \\
&= a \sum_x \sum_y x p_{X,Y}(x,y) + b \sum_x \sum_y y p_{X,Y}(x,y) \\
&= a \sum_x x p_{X,Y}(x,y) + b \sum_y y p_{X,Y}(x,y) \\
&= a\mathbb{E}[X] + b\mathbb{E}[Y]
\end{aligned}
$$

□

Proof.

$$\mathbb{E}[X,Y] = \sum_x \sum_y xy p_{X,Y}(x,y)$$

As X and Y are independent

$$
\begin{aligned}
&= \sum_x \sum_y p_X(x) p_y(Y) \\
&= \sum_x x P_X(x) \sum_y y p_Y(y) \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}
$$

- If X and Y are independent, then $\mathbb{E}[X,Y] = \mathbb{E}[X]\,\mathbb{E}[Y]$ □
- For the independent variables X and Y, $\mathrm{var}(X+Y) = var(X) + var(Y)$

  Proof. Let $\mu = \mathbb{E}[X]$ and $\lambda = \mathbb{E}[Y]$

$$
\begin{aligned}
\mathrm{var}(X+Y) &= \mathbb{E}\left[(X+Y-\mu-\lambda)^2\right] \\
&= \mathbb{E}\left[\lambda^2 + 2\lambda\mu + \mu^2 + X^2 - 2\lambda X - 2\mu X + 2XY + Y^2 - 2\lambda Y - 2\mu Y\right] \\
&= \sum_x \sum_y \left(\lambda^2 + 2\lambda\mu + \mu^2 + X^2 - 2\lambda X - 2\mu X + 2XY + Y^2 - 2\lambda Y - 2\mu Y\right) p_{X,Y}(x,y) \\
&= \sum_x \left(X^2 - 2\lambda X - 2\mu X\right) p_{X,Y}(x,y) + \sum_x \sum_y 2XY p_{X,Y}(x,y) \\
&+ \sum_y \left(Y^2 - 2\lambda Y - 2\mu Y\right) p_{X,Y}(x,y) + \mu^2 + \lambda^2 + 2\lambda\mu \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 + \mathbb{E}[Y]^2 - \mathbb{E}[Y]^2 \\
&= \mathrm{var}(X) + \mathrm{var}(Y)
\end{aligned}
$$

□