Lecture notes MVE136 Fall 2019. Part 1: Probability

2. Discrete random variables

We carry out a *random experiment*. The random experiment has various possible *outcomes* ξ. The *sample space* S is the set of all possible outcomes ξ of the random experiment. An *event* A is a subset of the sample space $A \subset S$. A *probability measure* Pr assigns probabilities $\Pr(A)$ to all events $A \subset S$.

Definition 2.1. (Axioms for Probability Measures)

1. $\Pr(A) \geq 0$ for $A \subset S$,
2. $\Pr(S) = 1$,
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ for $A, B \subset S$ with $A \cap B = \emptyset$.

Corollary 2.2.

1. $\Pr(\overline{A}) = \Pr(A^c) = 1 - \Pr(A)$ for $A \subset S$,
2. $\Pr(\emptyset) = 0$,
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ for $A, B \subset S$.

Above $\overline{A}$ and A^c are two ways to denote *the complement set to* A, i.e., $S \setminus A$.

Proof. (1) $1 =_{[\text{Axiom 2}]} \Pr(S) = \Pr(A \cap A^c) =_{[\text{Axiom 3}]} \Pr(A) + \Pr(A^c)$.

(2) Take $A = S$ so that $A^c = \emptyset$ in (1).

(3) $\Pr(A \cup B) = \Pr(A \cup (B \cap A^c)) =_{[\text{Axiom 3}]} \Pr(A) + \Pr(A^c \cap B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ as $\Pr(B) = \Pr((A \cap B) \cup (A^c \cap B)) =_{[\text{Axiom 3}]} \Pr(A \cap B) + \Pr(A^c \cap B)$. □

Sometimes it is convenient to write $\Pr(A, B)$ for the probability $\Pr(A \cap B)$.

Example 2.1. For toss with two dice we have $S = \{(i,j) : i, j \in \{1, \ldots, 6\}\}$. For fair dice $\Pr(\{(i,j)\}) = \frac{1}{36}$ for every (i,j) so that $\Pr(A) = \sum_{(i,j)\in A} \Pr(\{(i,j)\}) = \frac{\#\{(i,j)\in S : (i,j)\in A\}}{36}$ for $A \subset S$. For unfair dice any probability measure Pr that complies with Axioms 1-3 is possible but then the second of the previous equalities is no longer valid. An example of an event is $A = \{\text{sum of dice} \geq 10\} = \{(4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}$ with $\Pr(A) = \frac{1}{6}$ (for fair dice).

Definition 2.3. The *conditional probability* $\Pr(A|B)$ for $A \subset S$ given that $B \subset S$ occures (occured) is defined $\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)}$ [provided that $\Pr(B) > 0$].

Think of a deal at a poker table where you happen to see that one of your competators at the table obtained ace of spades. Then the conditional probability for four of a kind of ace in your own poker hand given that event is zero.

Example 2.1. (CONTINUED) If B is the event that first dice is 5 and A is as before we have $\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)} = \frac{\Pr(\{(5,5),(5,6)\})}{1/6} = \frac{2/36}{1/6} = \frac{2}{6} > \frac{1}{6} = \Pr(A)$.

Theorem 2.4. (LAW OF TOTAL PROBABILITY) For events A and mutually disjoint events $B_1, \ldots, B_n$ with $\cup_{i=1}^n B_i = S$ we have $\Pr(A) = \sum_{i=1}^n \Pr(A|B_i) \Pr(B_i)$.

The events $B_1, \ldots, B_n$ are called mutually disjoint when $B_i \cap B_j = \emptyset$ for $i \neq j$.

Proof. $\sum_{i=1}^n \Pr(A|B_i) \Pr(B_i) = \sum_{i=1}^n \frac{\Pr(A,B_i)}{\Pr(B_i)} \Pr(B_i) = \sum_{i=1}^n \Pr(A \cap B_i) =_{[\text{Axiom 3}]} = \Pr(A)$ since $A \cap B_1, \ldots, A \cap B_n$ are mutually disjoint with $\cup_{i=1}^n A \cap B_i = A$. □

Corollary 2.5. (BAYES' THEOREM) For events $A, B_1, \ldots, B_n$ as in the previous theorem we have $\Pr(B_j|A) = \frac{\Pr(A|B_j) \Pr(B_j)}{\sum_{i=1}^n \Pr(A|B_i) \Pr(B_i)}$.

Proof. The numerator of the right hand side is $\Pr(A \cap B_j)$ by the definition of conditional probability while the denominator is $\Pr(A)$ by the law of total probability. □

Definition 2.6. Two events A and B are independent if $\Pr(A \cap B) = \Pr(A) \Pr(B)$. The events $A_1, \ldots, A_n$ are indepedent if $\Pr(A_{i_1} \cap \ldots \cap A_{i_k}) = \Pr(A_{i_1}) \cdot \ldots \cdot \Pr(A_{i_k})$ whenever $i_1, \ldots, i_k \in \{1, \ldots, n\}$ are distinct.

Corollary 2.7. For A and B independent we have $\Pr(A|B) = \Pr(A)$.

Proof. For A and B independent $\Pr(A|B) = \frac{\Pr(A,B)}{\Pr(B)} = \frac{\Pr(A) \Pr(B)}{\Pr(B)} = \Pr(A)$. □

Example 2.1. (CONTINUED) A and B are not independent as $\Pr(A|B) \neq \Pr(B)$.

Definition 2.8. A *random variable (r.v.)* $X(\xi)$, $\xi \in S$, is a function $X : S \to \mathbb{R}$ from a sample space to the real numbers.

So a random variable is a function of the outcome of a random experiment. To find the value of the random variable one has to carry out the random experiment. Although $X(\xi)$ is a function of the outcome $\xi \in S$ of the experiment the dependence on ξ is virtually never indicated in the notation so that one writes just X instead of $X(\xi)$.

Definition 2.9. An r.v. is *discrete* if its number of different possible values is finite or countably infinite. Otherwise it is *continuous*.

Definition 2.10. The *probability mass function (PMF)* P_X for a discrete r.v. is defined $P_X(k) = \Pr(X = k)$ for all k.

Above $\Pr(X = x)$ is short hand notation for $\Pr(\{\xi \in S : X(\xi) = x\})$ - remember that Pr assigns probabilities to events and $\{X = x\}$ is short hand notation for the event $\{\xi \in S : X(\xi) = x\}$ that the outcome ξ of the random experiment is such that $X(\xi) = x$.

Example 2.1. (Continued) For X the sum of two dice $P_X(k) = \frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}$ and $\frac{1}{36}$ for $k = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11$ and 12, respectively.

Theorem 2.11. (Properties of PMF)

1. $\sum_{\text{all } k} P_X(k) = 1$,
2. $\Pr(X \in A) = \sum_{k \in A} P_X(k)$ for $A \subset \mathbb{R}$.

Proof. Clearly, (1) follows from (2) taking $A = \{\text{all possible values of } X\}$. Further, $\Pr(X \in A) = \Pr(\cup_{x \in A}\{X = x\}) =_{[\text{Axiom 3}]} \sum_{x \in A} \Pr(\{X = x\}) = \sum_{x \in A} P_X(x)$. □

Out of many namned discrete r.v.'s the following four argubly are most important:

Definition 2.12. A *Bernoulli r.v.* X has possible values $\{0, 1\}$ with PMF $P_X(0) = 1-p$ and $P_X(1) = p$ for a constant $p \in [0, 1]$.

One can interpret a Bernoulli r.v. as the indicator of whether one get heads $X = 1$ or tails $X = 0$ in a coin tossing experiment (of a not necessarily fair/balanced coin).

Definition 2.13. A *binomial r.v.* X has possible values $\{0, 1, \ldots, n\}$ with PMF $P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \ldots, n$ for constants $p \in [0, 1]$ and $n \in \mathbb{N}$.

One can interpret a binomial r.v. as the number of heads k one gets when tossing a coin n times (=the sum of n Bernoulli r.v.'s) since the probability of each the $\binom{n}{k}$ distinct ordered sequences of k heads and $n-k$ tails has probability $p^k(1-p)^{n-k}$.

Definition 2.14. A *Poisson r.v.* X has possible values $\mathbb{N}$ with PMF $P_X(k) = \alpha^k \mathrm{e}^{-\alpha}/(k!)$ for $k \in \mathbb{N}$ for a constant $\alpha > 0$.

Poisson r.v.'s occur naturally as, e.g., the number of radioactive decays per time unit of a piece of radioactive matter.

Definition 2.15. A *waiting time r.v.* X has possible values $\{1,2,\ldots\}$ with PMF $P_X(k) = p\,(1-p)^{k-1}$ for $k = 1,2,\ldots$ for a constant $p \in [0,1]$.

Waiting time r.v.'s can also be called *geometric r.v.'s*. One can interpret a waiting time r.v. as the number of tosses k of a coin that are needed to obtain the first heads.

3. Continuous random variables

Definition 3.1. The *cummulative distribution function (CDF)* F_X for an r.v. X is defined $F_X(x) = \Pr(X \le x)$ for $x \in \mathbb{R}$.

Theorem 3.2. (Properties of CDF)

1. $F_X(x) \in [0,1]$ with $F_X(-\infty) = 0$ and $F_X(\infty) = 1$,
2. $F_X(\underline{x}) \le F_X(\overline{x})$ for $\underline{x} \le \overline{x}$,
3. $\Pr(\underline{x} < X \le \overline{x}) = F_X(\overline{x}) - F_X(\underline{x})$.

Proof. By inspection. □

Example 2.1. (Continued) For X the sum of two dice we have $F_X(x) = 0, \frac{1}{36}, \frac{3}{36}, \frac{6}{36}, \frac{10}{36}, \frac{15}{36}, \frac{21}{36}, \frac{26}{36}, \frac{30}{36}, \frac{33}{36}, \frac{35}{36}$ and 1 for $x \in (-\infty,2)$, $[2,3)$, $[3,4)$, $[4,5)$, $[5,6)$, $[6,7)$, $[7,8)$, $[8,9)$, $[9,10)$, $[10,11)$, $[11,12)$ and $[12,\infty)$, respectively.

Definition 3.3. The *probability density function (PDF)* f_X for a continuous r.v. X is defined $f_X(x) = F_X'(x)$ for $x \in \mathbb{R}$.

Theorem 3.4. (Properties of PDF)

1. $f_X(x) \ge 0$ and $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$,
2. $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$,
3. $\Pr(X \in A) = \int_{x\in A} f_X(x)\,dx$ for $A \subset \mathbb{R}$.

Proof. (1) The integral formula follows from taking $A = \mathbb{R}$ in (3).

(2) Follows using that $F_X(x)$ is primitive function to $f_X(x)$ with $F_X(-\infty) = 0$.

(3) All sets $A \subseteq \mathbb{R}$ can be built by sets of type $(-\infty, x]$ using basic set operations and therefore (3) can be derived from (2) using the axioms of probability measures. □

Out of many namned continuous r.v.'s the following three are most important:

Definition 3.5. A *uniform r.v. X over the interval* $[a,b]$ has possible values $[a,b]$ with PDF $f_X(x) = 1/(a-b)$ for constants $-\infty < a < b < +\infty$.

Above we use the convention that $f_X(x) = 0$ in those regions where its values is not specified. A uniform r.v. over $[a,b]$ takes values in $[a,b]$ without any preference to have a more probable value in a certain sub-region of that interval.

A uniform r.v. X over $[a,b]$ has CDF $F_X(x) = (x-a)/(b-a)$ for $x \in [a,b]$. Here we use the convention that $F_X(x) = 0$ or 1 in those regions where its values is not specified.

Definition 3.6. An *exponential r.v.* X has possible values $\mathbb{R}^+$ with PDF $f_X(x) = \mathrm{e}^{-x/b}/b$ for $x \geq 0$ for a constant $b > 0$.

Some sources (e.g., Mathematica) use the parametrization $f_X(x) = b\,\mathrm{e}^{-bx}$ so that one must check what parametrization is in use. Exponential r.v.'s occur naturally as the time between two consecuitve radioactive decays of a piece of radioactive matter.

An exponential r.v. X has CDF $F_X(x) = 1 - \mathrm{e}^{-x/b}$ for $x \geq 0$.

Definition 3.7. A *Gaussian/Normal/*$\mathrm{N}(m,\sigma^2)$ r.v. X has possible values $\mathbb{R}$ with PDF $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\,\mathrm{e}^{-(x-m)^2/(2\,\sigma^2)}$ for $x \in \mathbb{R}$ for $m \in \mathbb{R}$ and $\sigma > 0$ constants.

Some sources (e.g., Matlab) use the parametrization $\mathrm{N}(m,\sigma)$ so that one must check what parametrization is in use. Gaussian random variables occur naturally because of the so called central limit theorem according to which the macroscopic sum of a very large number of independent equally distributed microscopic contributions is Gaussian.

Definition 3.8. An $\mathrm{N}(0,1)$ r.v. is called *standardized Gaussian* and its PDF and CDF are denoted $\phi(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2}$ and $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-y^2}\,dy$, respectively.

Theorem 3.9. For a general $\mathrm{N}(m,\sigma^2)$ r.v. X we have $F_X(x) = \Phi\big(\frac{x-m}{\sigma}\big)$.

Proof.

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-(x-m)^2/(2\,\sigma^2)}\,dy =_{[z=(y-m)/\sigma]} = \int_{-\infty}^{(x-m)/\sigma} \frac{1}{\sqrt{2\pi}}\,e^{-z^2/2}\,dz$$
$$= \Phi\Big(\frac{x-m}{\sigma}\Big). \qquad \square$$

Definition 3.10. The *conditional CDF* $F_{X|A}$ of an r.v. X given that an event $A \subset S$ occures (occured) is defined $F_{X|A}(x) = \frac{\Pr(X \le x, A)}{\Pr(A)}$ for $x \in \mathbb{R}$.

A conditional CDF $F_{X|A}$ has the same properties (1)-(3) as an unconditional CDF.

Definition 3.11. The *conditional PMF* $P_{X|A}$ of a discrete r.v. X given that an event $A \subset S$ occures (occured) is defined by $P_{X|A}(k) = F'_{X|A}(x)$ for $x \in \mathbb{R}$.

The *conditional PDF* $f_{X|A}$ of a continuous r.v. X given that an event $A \subset S$ occures (occured) is defined by $f_{X|A}(x) = F'_{X|A}(x)$ for $x \in \mathbb{R}$.

A conditional PMF $P_{X|A}$ has same properties (1)-(2) as an unconditional PMF and a conditional PDF $f_{X|A} =$ has same properties (1)-(3) as an unconditional PDF.

Of special interest is the conditional CDF and PDF of a continuous r.v. X given that the event $A = \{a < X \le b\}$ occures (occured) in which case one readily gets

$$F_{X|A}(x) = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} \quad \text{and} \quad f_{X|A}(x) = \frac{f_X(x)}{F_X(a) - F_X(b)} \quad \text{for } a < x \le b.$$

4. Operations on random variables

Definition 4.1. The *expected value* $E(X) = \mu_X$ of an r.v. X is defined

$$E(X) = \begin{cases} \sum_{\text{all } k} k\, P_X(k) & \text{for } X \text{ discrete,} \\ \int_{-\infty}^{\infty} x\, f_X(x)\, dx & \text{for } X \text{ continuous.} \end{cases}$$

The expected value is the center of gravity for the PMF or PDF.

Example 2.1. (Continued) For X the sum of two dice we have $E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot 436 + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = \ldots = 7$.

Theorem 4.2. For a function $g : \mathbb{R} \to \mathbb{R}$ we have

$$E(g(X)) = \begin{cases} \sum_{\text{all } k} g(k)\, P_X(k) & \text{for } X \text{ discrete,} \\ \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & \text{for } X \text{ continuous.} \end{cases}$$

Proof. (Discrete case.)

$$E(g(X)) = \sum_{\text{all } k} k \Pr(g(X) = k) =_{[k=g(\ell)]} \sum_{\text{all } \ell} g(\ell) \Pr(g(X) = g(\ell)) = \sum_{\text{all } \ell} g(\ell)\, P_X(\ell). \quad \square$$

Definition 4.3. The *n'th moment of an r.v.* X is defined $E(X^n)$ for $n \in \mathbb{N}$. The *n'th central moment of an r.v.* X is defined $E((X - \mu_X)^n)$ for $n \in \mathbb{N}$.

Example 4.1. For X uniformly distributed over $[a, b]$ we have

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_a^b \tfrac{x}{b-a}\, dx = \left[\tfrac{x^2}{2\,(b-a)}\right]_{x=a}^{x=b} = \tfrac{b^2-a^2}{2\,(b-a)} = \tfrac{a+b}{2}.$$

Definition 4.4. The *conditional expected value* $E(X|A)$ of an r.v. X given that an event $A \subset S$ occures (occured) is defined

$$E(X|A) = \begin{cases} \sum_{\text{all } k} k\, P_{X|A}(k) & \text{for } X \text{ discrete,} \\ \int_{-\infty}^{\infty} x\, f_{X|A}(x)\, dx & \text{for } X \text{ continuous.} \end{cases}$$

Theorem 4.5. For X a continuous r.v. with PDF $f_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ strictly increasing the r.v. $Y = g(X)$ has PDF $f_Y(y) = f_X(g^{-1}(y))\, \frac{d}{dy}\, g^{-1}(y)$.

Proof. $f_Y(y) = \frac{d}{dy} \Pr(g(X) \le y) = \frac{d}{dy} \Pr(X \le g^{-1}(y)) = f_X(g^{-1}(y))\, \frac{d}{dy}\, g^{-1}(y)$. □

Example 4.1. (Continued) For X uniformly distributed over $[0, 1]$ with $f_X(x) = 1$ for $x \in [0, 1]$ and $Y = -\ln(X)$ we cannot apply the previous theorem since $-\ln(x)$ is not increasing, but by direct calculation we get

$$\begin{aligned} f_Y(y) = \tfrac{d}{dy} \Pr(Y \le y) &= \tfrac{d}{dy} \Pr(-\ln(X) \le y) = \tfrac{d}{dy} \Pr(X \ge \mathrm{e}^{-y}) \\ &= \tfrac{d}{dy}\big(1 - \Pr(X < \mathrm{e}^{-y})\big) = \mathrm{e}^{-y}\, f_X(\mathrm{e}^{-y}) = \mathrm{e}^{-y} \end{aligned}$$

for $y \ge 0$ so that Y is exponentially distributed with parameter 1.

Definition 4.6. The *characteristic function (CHF)* Φ_X of an r.v. X is defined $\Phi_X(\omega) = E(\mathrm{e}^{j\omega X})$ for $\omega \in \mathbb{R}$ (where j is the imaginary unit $j^2 = -1$).

Here $\Phi_X(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{j\omega x} f_X(x)\, dx$ is the Fourier transform of f_X for X continuous.

Example 4.2. For X an $\mathrm{N}(m, \sigma^2)$ distributed r.v. we have

$$\begin{aligned} \Phi_X(\omega) &= \int_{-\infty}^{\infty} \mathrm{e}^{j\omega x}\, \frac{1}{\sqrt{2\pi}\,\sigma}\, \mathrm{e}^{-(x-m)^2/(2\,\sigma^2)}\, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, \mathrm{e}^{-(x-m-j\omega\sigma^2)^2/(2\,\sigma^2)}\, dx \times \mathrm{e}^{j\omega m - \frac{1}{2}\omega^2\sigma^2} \\ &= \int_{-\infty}^{\infty} f_{\mathrm{N}(m+j\omega\sigma^2, \sigma^2)}(x)\, dx \times \mathrm{e}^{j\omega m - \frac{1}{2}\omega^2\sigma^2} = \mathrm{e}^{jm\omega - \frac{1}{2}\sigma^2\omega^2}. \end{aligned}$$

In the above example we used the useful trick to recognize an integral of a PDF (or the sum of a PMF) to conclude that the integral (sum) is one without calculations.

Theorem 4.7. $(-j)^n\,\Phi_X^{(n)}(0) = E(X^n)$ for $n \in \mathbb{N}$.

Proof. $\Phi_X^{(n)}(0) = \frac{d^n}{d\omega^n} E(\mathrm{e}^{j\omega X})\big|_{\omega=0} = E(\frac{d^n}{d\omega^n}\mathrm{e}^{j\omega X})\big|_{\omega=0} = E((jX)^n \mathrm{e}^{j\omega X})\big|_{\omega=0} = j^n E(X^n)$. □

Example 4.2. (Continued) For X an $\mathrm{N}(m,\sigma^2)$ distributed r.v. we have

$$E(X) = (-j)\,\Phi_X'(0) = (-j)\,(jm-\sigma^2\omega)\,\Phi_X(\omega)\big|_{\omega=0} = m,$$

$$E(X^2) = (-j)^2\Phi_X''(0) = j^2\big((jm-\sigma^2\omega)^2-\sigma^2\big)\,\Phi_X(\omega)\big|_{\omega=0} = m^2+\sigma^2.$$

The moments of the six other namned random variables we have considered so far can also be calculated by differentiating their CHF at zero.

Definition 4.8. The *probability generating function (PGF)* H_X of an $\mathbb{N}$-valued r.v. X is defined $H_X(z) = E(z^X) = \sum_{k=0}^{\infty} z^k\, P_X(k)$ for $z \in [0,1]$.

Theorem 4.9. $H_X^{(n)}(1) = E\big(X\,(X-1)\cdot\ldots\cdot(X-n+1)\big)$ and $P_X(n) = \frac{1}{n!}\,H_X^{(n)}(0)$ for $n \in \mathbb{N}$.

Proof. $\frac{d^n}{dz^n}E(z^X)\big|_{z=1} = E(\frac{d^n}{dz^n}z^X)\big|_{z=1} = E\big(X\,(X-1)\cdot\ldots\cdot(X-n+1)\,z^X\big)\big|_{z=1} = E\big(X\,(X-1)\cdot\ldots\cdot(X-n+1)\big)$ while the formula for $P_X(n)$ is Taylor expansion. □

5. Two-dimensional random variables

Definition 5.1. The *joint CDF* $F_{X,Y}$ for a pair of r.v.'s (X,Y) is defined $F_{X,Y}(x,y) = \Pr(X \le x,\, Y \le y)$ for $x, y \in \mathbb{R}$.

Theorem 5.2. (Properties of Joint CDF)

1. $F_{X,Y}(x,y) \in [0,1]$, $F_{X,Y}(-\infty,y) = F_{X,Y}(x,-\infty) = 0$ and $F_{X,Y}(\infty,\infty) = 1$,
2. $F_{X,Y}(x,\infty) = F_X(x)$ and $F_{X,Y}(\infty,y) = F_Y(y)$
3. $F_{X,Y}(\underline{x},y) \le F_{X,Y}(\overline{x},y)$ for $\underline{x} \le \overline{x}$ and $F_{X,Y}(x,\underline{y}) \le F_{X,Y}(x,\overline{y})$ for $\underline{y} \le \overline{y}$,
4. $\Pr(\underline{x} < X \le \overline{x},\, \underline{y} < Y \le \overline{y}) = F_{X,Y}(\overline{x},\overline{y}) - F_{X,Y}(\overline{x},\underline{y}) - F_{X,Y}(\underline{x},\overline{y}) + F_{X,Y}(\underline{x},\underline{y})$.

Proof. (1)-(3) are by inspection while (4) follows subtracting $\Pr(X \le \underline{x}, \underline{y} < Y \le \overline{y}) = F_{X,Y}(\underline{x},\overline{y}) - F_{X,Y}(\underline{x},\underline{y})$ from $\Pr(X \le \overline{x}, \underline{y} < Y \le \overline{y}) = F_{X,Y}(\overline{x},\overline{y}) - F_{X,Y}(\overline{x},\underline{y})$. □

The definition of (X,Y) being *discrete* and *continuous* , respectively, is as before.

Definition 5.3. The *joint PMF* $P_{X,Y}$ for a pair of discrete r.v.'s (X,Y) is defined $P_{X,Y}(k,\ell) = \Pr(X = k, Y = \ell)$ for all k, ℓ.

Theorem 5.4. (Properties of Joint PMF)

1. $P_{X,Y}(k,\ell) \in [0,1]$ and $\sum\sum_{\text{all } k,\ell} P_{X,Y}(k,\ell) = 1$,
2. $P_X(k) = \sum_{\text{all } \ell} P_{X,Y}(k,\ell)$ and $P_Y(\ell) = \sum_{\text{all } k} P_{X,Y}(k,\ell)$,
3. $\Pr((X,Y) \in A) = \sum\sum_{\text{all } (k,\ell)\in A} P_{X,Y}(k,\ell)$ for $A \subset \mathbb{R}^2$.

Proof. By analogy with the proof for one dimensional r.v. □

Definition 5.5. The *joint PDF* $f_{X,Y}$ for a pair continuous r.v.'s (X,Y) is defined $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$ for $x, y \in \mathbb{R}$.

Theorem 5.6. (Properties of Joint PDF)

1. $f_{X,Y}(x,y) \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dxdy = 1$,
2. $F_{X,Y}(x,y) = \int_{u=-\infty}^{u=x}\int_{v=-\infty}^{v=y} f_{X,Y}(u,v)\, dudv$,
3. $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$,
4. $\Pr((X,Y) \in A) = \iint_{\{(x,y)\in\mathbb{R}^2 \,:\, (x,y)\in A\}} f_{X,Y}(x,y)\, dxdy$ for $A \subset \mathbb{R}^2$.

Proof. (2) Follows as $F_{X,Y}$ is primitive to $f_{X,Y}$ with $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$.

(3) Follows by differentiation of (2).

(4) All $A \subseteq \mathbb{R}^2$ can be built by sets of type $(-\infty, x] \times (-\infty, y]$ using basic set operations and thus (4) can be derived from (2) using the axioms of probability measures.

(1) Follows since $f_{X,Y}(x,y) \geq 0$ by (4) and taking $A = \mathrm{R}^2$ in (4), respectively. □

Example 5.1. For $f_{X,Y}(x,y) = \frac{1}{2}\,\mathrm{e}^{-x-y/2}$ for $x, y \geq 0$ we have

$$\Pr(X > Y) = \iint_{\{(x,y)\in\mathbb{R}^2 \,:\, x>y\}} f_{X,Y}(x,y)\, dxdy = \int_0^\infty \left(\int_y^\infty \mathrm{e}^{-x}\, dx\right) \tfrac{1}{2}\,\mathrm{e}^{-y/2}\, dy = \frac{1}{3}.$$

Example 5.2. For $f_{X,Y}(x,y) = \frac{1}{\sqrt{3}\,\pi}\,\mathrm{e}^{-\frac{2}{3}(x^2-xy+y^2)}$ for $x, y \in \mathbb{R}$ we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{3/4}}\,\mathrm{e}^{(y-\frac{1}{2}x)^2/(2\,(3/4))}\, dx\, \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2} = \phi(x),$$

since the integrand of the second integral is $f_{\mathrm{N}(x/2,3/4)}(y)$ so the integral is 1.

Theorem 5.7. For a function $g : \mathbb{R}^2 \to \mathbb{R}$ we have

$$E(g(X,Y)) = \begin{cases} \sum\sum_{\text{all } k,\ell} g(k,\ell)\, P_{X,Y}(k,\ell) & \text{for } (X,Y) \text{ discrete,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dxdy & \text{for } (X,Y) \text{ continuous.} \end{cases}$$

Proof. By analogy with the proof for one dimensional r.v. □

Theorem 5.8. (Linearity) For r.v.'s $X_1, \dots, X_n$ and $a_1, \dots, a_n \in \mathbb{R}$ we have

$$E\big(\textstyle\sum_{i=1}^n a_i X_i\big) = \sum_{i=1}^n a_i E(X_i).$$

Proof. By property (3) of joint PDF we have

$$\begin{aligned} E\big(\textstyle\sum_{i=1}^n a_i X_i\big) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \big(\textstyle\sum_{i=1}^n a_i x_i\big) f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 \dots dx_n \\ &= \textstyle\sum_{i=1}^n a_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 \dots dx_n \\ &= \textstyle\sum_{i=1}^n a_i \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)\, dx_i = \sum_{i=1}^n a_i E(X_i). \end{aligned}$$ □

Definition 5.9. The *variance* $\mathrm{Var}(X) = \sigma_X^2$ an r.v. X is defined $\mathrm{Var}(X) = E((X-\mu_X)^2)$. The *standarddeviation* of X is $\sigma_X = \sqrt{\mathrm{Var}(X)}$.

Definition 5.10. The *correlation* $R_{X,Y}$ between two r.v.'s X and Y is defined $R_{X,Y} = E(XY)$. The *covariance* $\mathrm{Cov}(Y,Y)$ between X and Y is defined $\mathrm{Cov}(Y,Y) = E((X-\mu_X)(y-\mu_Y))$.

Clearly $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$ and from linearity of the mean we see that $\mathrm{Cov}(X,Y) = R_{X,Y} - \mu_X\mu_Y$. More generally we obtain the following result in the same fashion:

Theorem 5.11.

1. $R_{\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j} = \sum_{i=1}^m \sum_{j=1}^n a_i b_j\, R_{X_i,Y_j}$,
2. $\mathrm{Cov}\big(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\big) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j\, \mathrm{Cov}(X_i, Y_j)$,
3. $\mathrm{Var}\big(\sum_{i=1}^m a_i X_i\big) = \sum_{i=1}^m \sum_{j=1}^m a_i a_j\, \mathrm{Cov}(X_i, X_j)$.

Definition 5.12. The *correlation coefficient* $\rho_{X,Y}$ between X and Y is defined $\rho_{X,Y} = \mathrm{Cox}(X,Y)/\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$.

Theorem 5.13. $|\rho_{X<Y}| \leq 1$.

Proof. $0 \leq \mathrm{Var}\big((X/\sigma_X) \pm (Y/\sigma_Y)\big) = 2 \pm 2\,\rho_{X,Y}$. □

Definition 5.14. Two r.v.'s X and Y are *independent* if $\Pr(X \in A, Y \in B) = \Pr(X \in A)\Pr(Y \in B)$ for $A, B \subset \mathbb{R}$.

Theorem 5.15. Two r.v.'s X and Y are independent if and only if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ if and only if $P_{X,Y}(k,\ell) = P_X(k)P_Y(\ell)$ for (X,Y) discrete or $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for (X,Y) continuous.

Proof. By inspection. □

Definition 5.16. Two r.v.'s X and Y are *uncorrelated* if $\mathrm{Cov}(X,Y) = 0$.

Theorem 5.17. Two independent r.v.'s X and Y are uncorrelated.

Proof. (Continuous case.) As $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for X and Y independent

$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{X,Y}(x,y)\,dxdy = \big(\int_{-\infty}^{\infty} x\, f_X(x)\,dx\big)\,\big(\int_{-\infty}^{\infty} y\, f_Y(y)\,dy\big) = \mu_X\mu_Y$. □

Definition 5.18. A *jointly Gaussian/Normal* r.v. (X,Y) has PDF

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}}\,\exp\left\{-\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{X,Y}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{2\,(1-\rho_{X,Y}^2)}\right\}$$

for $(x,y) \in \mathbb{R}^2$.

In the above definition it is readily shown by direct calculation that $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ and $\rho_{X,Y}$ are the expectations, variances and correlation coefficient for X and Y.

Example 5.2. (Continued) The PDF $f_{X,Y}(x,y) = \frac{1}{\sqrt{3}\,\pi}\,\mathrm{e}^{-\frac{2}{3}(x^2-xy+y^2)}$ is bivariate Gaussian with $\mu_X = \mu_Y = 0$, $\sigma_X^2 = \sigma_Y^2 = 1$ and $\rho_{X,Y} = \frac{1}{2}$.

Definition 5.19. For a discrete pair of r.v.'s (X,Y) he conditional PMF of X given that $Y = y$ is defined $P_{X|Y}(x|y) = \Pr(X = x|Y = y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$.

For a continuous pair of r.v.'s (X,Y) the conditional PDF of X given that $Y = y$ is defined $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$.

Definition 5.20.

$$\Pr(X \in A|Y = y) = \begin{cases} \sum_{\text{all } x\in A} P_{X|Y}(x|y) & \text{for } (X,Y) \text{ discrete,} \\ \int_{x\in A} f_{X|Y}(x|y)\,dx & \text{for } (X,Y) \text{ continuous.} \end{cases}$$

Definition 5.21.

$$E(X|Y = y) = \begin{cases} \sum_{\text{all } x} x\, P_{X|Y}(x|y) & \text{for } (X,Y) \text{ discrete,} \\ \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\,dx & \text{for } (X,Y) \text{ continuous.} \end{cases}$$

Theorem 5.22. (Law of Total Probability)

$$\Pr(X \in A) = \begin{cases} \sum_{\text{all } y} \Pr(X \in A | Y = y)\, P_Y(y) & \text{for } (X, Y) \text{ discrete,} \\ \int_{-\infty}^{\infty} \Pr(X \in A | Y = y)\, f_Y(y)\, dy & \text{for } (X, Y) \text{ continuous.} \end{cases}$$

Proof. (Discrete case.) $\sum_{\text{all } y} \Pr(X \in A | Y = y)\, P_Y(y) = \sum \sum_{\text{all } x \in A, \text{ all } y} P_{X,Y}(x, y)$. $\square$

In an entirely similar fashion one proves the following result:

Theorem 5.23. (Law of Total Expectation)

$$E(X) = \begin{cases} \sum_{\text{all } x} E(X | Y = y)\, P_Y(y) & \text{for } (X, Y) \text{ discrete,} \\ \int_{-\infty}^{\infty} E(X | Y = y)\, f_Y(y)\, dy & \text{for } (X, Y) \text{ continuous.} \end{cases}$$

Definition 5.24. The *joint moment* of an r.v. (X, Y) is defined $E(X^m Y^n)$ and the *joint central moment* $E\big((X - \mu_X)^m (Y - \mu_Y)^n\big)$ for $m, n \in \mathbb{N}$.

Definition 5.25. The *joint CHF* $\Phi_{X,Y}$ for a pair of r.v. (X, Y) is defined $\Phi_{X,Y}(\omega_1, \omega_2) = E(\mathrm{e}^{j(\omega_1 X + \omega_2 Y)})$ for $\omega_1, \omega_2 \in \mathbb{R}$.

Definition 5.26. The *joint PGF* $H_{X,Y}$ of an $\mathbb{N}^2$-valued r.v. (X, Y) is defined $H_{X,Y}(z_1, z_2) = E(z_1^X z_2^Y) = \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} z_1^k z_2^\ell P_{X,Y}(k, \ell)$ for $z_1, z_2 \in [0, 1]$.

By analogy with the one dimensional case we obtain the following two results:

Theorem 5.27. $(-j)^{m+n}\, \Phi_{X,Y}^{(m,n)}(0, 0) = E(X^m Y^n)$ for $m, n \in \mathbb{N}$.

Theorem 5.28. $H_{X,Y}^{(m,n)}(1, 1) = E\big(X\,(X - 1) \cdot \ldots \cdot (X - m + 1)\, Y\,(Y - 1) \cdot \ldots \cdot (Y - n + 1)\big)$ and $P_{X,Y}(m, n) = \frac{1}{m!\,n!}\, H_{X,Y}^{(m,n)}(0, 0)$ for $m, n \in \mathbb{N}$.

Theorem 5.29. If X and Y are independent r.v.'s then

$$\begin{cases} P_{X+Y}(m) = \sum_{\text{all } k} P_X(k) P_Y(m-k) = \sum_{\text{all } \ell} P_X(m-\ell) P_Y(\ell) & \text{for } X \text{ and } Y \text{ discrete,} \\ f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy & \text{for } X \text{ and } Y \text{ continuous.} \end{cases}$$

Proof. (Continuous case.)

$$\begin{aligned} f_{X+Y}(z) = \frac{d}{dz} \Pr(X+Y \le z) &= \frac{d}{dz} \iint_{\{(x,y) \in \mathbb{R}^2 \,:\, x+y \le z\}} f_{X,Y}(x, y)\, dx dy \\ &= \frac{d}{dz} \int_{-\infty}^{\infty} f_X(x) \Big(\int_{-\infty}^{z-x} f_Y(y)\, dy \Big) dx = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx. \end{aligned}$$

The proof can alternatively be done using CHF techniques as in the next example. □

Example 4.2. (Continued) For X and Y independent $\mathrm{N}(m_1, \sigma_1^2)$ and $\mathrm{N}(m_2, \sigma_2^2)$ distributed, respectively, we have

$$\Phi_{X+Y}(\omega) = E(\mathrm{e}^{j\omega(X+Y)}) = E(\mathrm{e}^{j\omega X}\mathrm{e}^{j\omega Y}) = E(\mathrm{e}^{j\omega X})\, E(\mathrm{e}^{j\omega Y}) = \mathrm{e}^{j(m_1+m_2)\omega - \frac{1}{2}(\sigma_1^2+\sigma_2^2)\omega^2},$$

since $\mathrm{e}^{j\omega X}$ and $\mathrm{e}^{j\omega Y}$ are uncorrelated. Hence $X+Y$ is $\mathrm{N}(m_1{+}m_2, \sigma_1^2{+}\sigma^2)$.

Theorem 5.30. For a continuous r.v. (X, Y) with joint PDF $f_{X,Y}$ and functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$ the r.v. $(Z, W) = (g_1(X), g_2(Y))$ is continuous with

$$f_{Z,W}(z, w) = f_{X,Y}(x, y) \Big/ \left\| \begin{array}{cc} \partial z/\partial x & \partial z/\partial y \\ \partial w/\partial x & \partial w/\partial y \end{array} \right\| \Bigg|_{(x,y)=(h_1(z,w),h_2(z,w))},$$

where $(h_1, h_2) : \mathbb{R}^2 \to \mathbb{R}^2$ is the inverse transformation to $(g_1, g_2) : \mathbb{R}^2 \to \mathbb{R}^2$.

Proof. By change of variable in two dimensional integral we have

$$\Pr((Z, W) \in A) = \Pr\big((g_1(X, Y), g_2(X, Y)) \in A\big)$$

$$= \iint_{\{(x,y)\in\mathbb{R}^2 \,:\, (g_1(x,y),g_2(x,y))\in A\}} f_{X,Y}(x, y)\, dxdy = \begin{bmatrix} x = h_1(z, w) \\ y = h_2(z, w) \end{bmatrix}$$

$$= \iint_{\{(x,y)\in\mathbb{R}^2 \,:\, (z,w)\in A\}} f_{X,Y}(h_1(z, w), h_2(z, w)) \Big/ \left\| \begin{array}{cc} \partial z/\partial x & \partial z/\partial y \\ \partial w/\partial x & \partial w/\partial y \end{array} \right\| \Bigg|_{\substack{x=h_1(z,w)\\ y=h_2(z,w)}} dzdw.$$

□