AMA563 Principle of Data Science

Chapter 1

Elementary Probability Theory

Contents

1.1 Probability space $(\mathcal{C}, \mathcal{B}, P)$

$\mathcal{C}$: Sample space (i.e. the set of all possible outcomes)

$\mathcal{B}$: Collection of events. Event: subset of $\mathcal{C}$

P: Probability. $P(A)$: probability of event A

(i) $P(A) \geq 0$ for $\forall A \in \mathcal{B}$

(ii) $P(\mathcal{C}) = 1$

(iii) $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ if $A_1, A_2, \ldots,$ are disjoint.

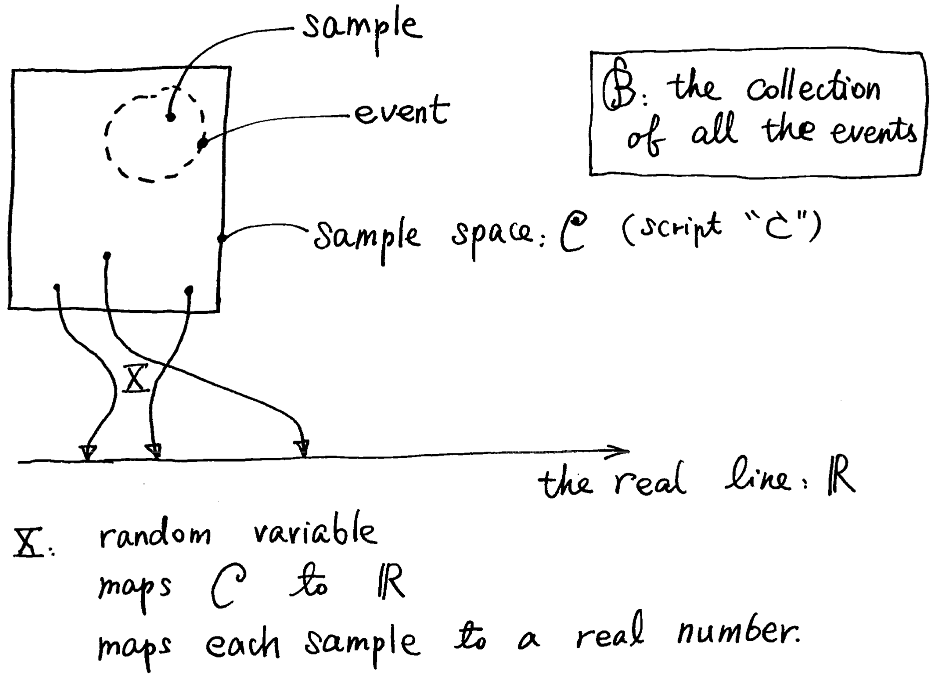
1.2 Random Variable (rv)

X : a function $\mathcal{C} \to \mathbb{R}$, where $\mathbb{R}$ is the set of all the real numbers

$\forall x \in \mathbb{R}$, $\{X \leq x\}$ is an event. Therefore the probability $P(X \leq x)$ is well defined.

$F(x) = P(X \leq x)$ is called a cumulative distribution function (cdf) of X.

About notation: We usually use lower case letters $(a, b, c, x, y, z, \ldots)$ to denote deterministic variables, and use upper case letters $(X, Y, Z, W, T, U, \ldots)$ to denote random variables. For example, "a random variable X may take value x or y with probability 1/2 each".



1.3 Type of rv's

We focus on two types of random variables: discrete rv's and continuous rv's. There are also mixed-type rv's which we do not expand here.

(1) **Discrete rv's:** the space or range $\mathcal{D}$ of X is countable (either finite or has as many elements as there are positive integers.[1])

Probability mass function (pmf): $p_X(x) = P(X = x),\ x \in \mathcal{D}$.

$0 \le p_X(x) \le 1$ and $\sum_{x\in\mathcal{D}} p_X(x) = 1$.

Transformations: Let X be a discrete random variable and $Y = g(X)$. Let p_X and p_Y be the pmf's of X and Y respectively. We have

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

In particular, when g is one-to-one, the only x such that $g(x) = y$ is $x = g^{-1}(y)$, so $\boxed{p_Y(y) = p_X(g^{-1}(y))}$

(2) **Continuous rv's:** $X \in (-\infty, \infty)$ or (a, b), $F(x)$ is differentiable

Probability density function (pdf): $f(x) = F'(x)$

$f(x) \ge 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$, $F(x) = \int_{-\infty}^{x} f(y)dy$

Transformations: $Y = g(X)$, g is a one-to-one differentiable function, $x = g^{-1}(y)$.

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|.$$

The general case when g is not one-to-one, is involved. For example when $g(x) \equiv 0$, then $g(X)$ is a degenerated random variable which takes only one value 0, with probability 1. However, it is easy to find the cumulative distribution function (cdf) of Y, $F_Y(y) = P(Y \le y) = P(g(X) \le y)$. If F_Y is differentiable, then

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(g(X) \le y).$$

[1]For example, the set $C_1 = \{1, 2, 3\}$ is finite, and thus countable; the set $C_2 = \{1, 2, 3, 4, 5, \ldots\}$ contains all the positive integers, and therefore countable. One needs more involved arguments to prove that the set $C_3 = \{0, \pm1, \pm2, \ldots\}$ is countable, and the set $C_4 = [0, 1]$ is not countable.

Example 1.3.1. *Let X have a pmf $p(x) = 1/3$, $x = 1, 2, 3$, zero elsewhere. Find the pmf of $Y = 2X + 1$. Find the pmf of $Z = X^2 + 1$.*

Solution.

x	1	2	3
$p_X(x)$	$1/3$	$1/3$	$1/3$
$2x+1$	3	5	7

y	3	5	7
$p_Y(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Therefore:

$$p_Y(y) = \begin{cases} \frac{1}{3}, & y = 3 \\ \frac{1}{3}, & y = 5 \\ \frac{1}{3}, & y = 7 \\ 0 & \text{otherwise.} \end{cases}$$

The pmf of $Z = X^2 + 1$ is left as an exercise.

Example 1.3.2. *The pdf of X is $f(x) = 2xe^{-x^2}$, $0 < x < \infty$, zero elsewhere. Determine the pdf of $Y = X^2$. Determine the pdf of $Z = \sqrt{X}$.*

Solution. Since $Y \geq 0$, we have that for all $y < 0$, $f_Y(0) = 0$. For any $y \geq 0$,

$$\begin{aligned} F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) &= \int_0^{\sqrt{y}} 2xe^{-x^2} dx \\ &= -e^{-x^2}\Big|_0^{\sqrt{y}} = 1 - e^{-y}. \\ f_Y(y) &= \frac{d}{dy} F_Y(y) = e^{-y}. \end{aligned}$$

Therefore

$$f_Y(y) = \begin{cases} e^{-y}, & y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The pdf for $Z = \sqrt{X}$ is left as an exercise.

1.4 Distributional quantities

(1) Mean or expectation of X

$$Eg(X) = \begin{cases} \sum_x xp(x), & \text{if } \sum |x|p(x) < \infty \quad \text{(discrete case)} \\ \int_{-\infty}^{\infty} xf(x)dx, & \text{if } \int |x|f(x)dx < \infty \quad \text{(continuous case)} \end{cases}$$

(2) Higher moments (m-th), where m is a positive integer

$$E(X^m) = \begin{cases} \sum_x x^m p(x), & \text{if } \sum |x|^m p(x) < \infty \quad \text{(discrete case)} \\ \int_{-\infty}^{\infty} x^m f(x)dx, & \text{if } \int |x|^m f(x)dx < \infty \quad \text{(continuous case)} \end{cases}$$

(3) Variance: $\sigma^2 = \mathrm{Var}(X) = E(X - E(X))^2 = EX^2 - (E(X))^2$

(4) Moment generating function (mgf)

If $E(e^{tX})$ exists for $|t| < h$, the mgf of X is defined to be the function $M(t) = E(e^{tX})$, $|t| < h$.

Properties:

- $M^{(m)}(0) = E(X^m)$, for every positive integer m.

- If X and Y are independent, $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Example 1.4.1. *We flip an unfair coin which gives a head with probability 0.4 and a tail with probability 0.6.*

Discussion:

In this example, $\mathcal{C} = \{\text{Head}, \text{Tail}\}$, which is the whole sample space. The event collection $\mathcal{B} = \{\emptyset, \{\text{Head}\}, \{\text{Tail}\}, \mathcal{C}\}$. Here,

- $\emptyset$ = the empty event. This is used to denote the event that is impossible. For example, $\{\text{Head}\} \cap \{\text{Tail}\} = \emptyset$, which means that we cannot get a head and a tail simultaneously. We have $P(\emptyset) = 0$.

- $\{\text{Head}\}$ and $\{\text{Tail}\}$ represent the events that we get a head and we get a tail, respectively. As we assumed, $P(\{\text{Head}\}) = 0.4$, and $P(\{\text{Tail}\}) = 0.6$.

- $\mathcal{C} = \{\text{Head}, \text{Tail}\}$ denotes the event that must happen. Here this event is that we get either a head, or a tail. Obviously this will definitely happen[2]. One has that $P(\mathcal{C}) = 1$.

Now we may define $X : \mathcal{C} \to \mathbb{R}$ as, for example,

$$X(\text{Head}) = 1, \qquad X(\text{Tail}) = -0.35.$$

Then X is a random variable which takes 1 with probability 0.4 and -0.35 with probability 0.6. People may use this random variable to model a bet, where one wins 1 dollar if the coin gives a head, and loses 0.35 dollar if the coin gives a tail. Obviously, X is a discrete rv. We have

$$E(X) = \sum_x xp(x) = 1 \times 0.4 + (-0.35) \times 0.6 = 0.19;$$
$$E(x^2) = 1^2 \times 0.4 + (-0.35)^2 \times 0.6 = 0.4735;$$
$$\cdots$$
$$E(X^m) = 1^m \times 0.4 + (-0.35)^m \times 0.6;$$
$$M_X(t) = E(e^{tX}) = e^t \times 0.4 + e^{-0.35t} \times 0.6, \quad -\infty < t < \infty.$$

1.5 Euler's integrals

Factorial: $n! = n \times (n-1) \times \cdots \times 1$ for any positive integer n. Also, we *define* $0! = 1$.

Gamma function: $\Gamma(x)$, defined on $x \in (0, \infty)$, by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Properties of $\Gamma(x)$:

- $\Gamma(n+1) = n!$ for $n = 0, 1, 2, \ldots$.

[2]OK, OK, I mean, we do not consider the case that the coin disappears, the coin stands up, etc., etc. Don't ask why.

- $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$.
- $\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}$ for $0 < x < 1$.
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Beta function: for $z > 0$ and $w > 0$,

$$\begin{aligned} B(z,w) &= \int_0^1 t^{z-1}(1-t)^{w-1}dt = \int_0^\infty \frac{t^{z-1}}{(1+t)^{z+w}}dt \\ &= 2\int_0^{\pi/2} (\sin t)^{2z-1}(\cos t)^{2w-1}dt = \frac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)}. \end{aligned}$$

Example 1.5.1. *Let $X \sim N(0,1)$. Find $E(X^{10})$ (the 10^{th} moment of X). Find also $E(X^5)$ and $E(X^6)$.*

Solution.

$$\begin{aligned} E(X^{10}) &= \int_{-\infty}^\infty x^{10}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \overset{y=x^2/2}{=\!=\!=} \sqrt{\frac{2}{\pi}}\int_0^\infty e^{-y}2^5y^5\frac{1}{\sqrt{2y}}dy \\ = \frac{2^5}{\sqrt{\pi}}\int_0^\infty e^{-y}y^{5.5-1}dy &= \frac{2^5}{\sqrt{\pi}}\Gamma(5.5) = \frac{2^5 \times 4.5 \times 3.5 \times 2.5 \times 1.5 \times 0.5\Gamma(0.5)}{\sqrt{\pi}} \\ &= 9 \times 7 \times 5 \times 3 \times 1 = 945. \end{aligned}$$

The moments $E(X^5)$ and $E(X^6)$ are left as exercises.

Example 1.5.2. *Let* $X \sim \mathsf{N}(0,1)$*. Find* $E(X^2)$ *and* $\operatorname{Var}(X^2)$*. Let* $Y_1, Y_2 \sim \mathsf{Exponential}(\lambda)$ *and assume they are independent. Find* $E(Y_1+Y_2)$*,* $\operatorname{Var}(Y_1+Y_2)$*,* $M_{2Y_1}(t)$*, and* $M_{Y_1+Y_2}(t)$*.*

Solution. We have

$$\begin{aligned}
E(X^2) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-x^2/2}dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^2 e^{-x^2/2}dx \\
\overset{y=\frac{x^2}{2}}{\underset{x=\sqrt{2y}}{=\!=\!=}}\ & \sqrt{\frac{2}{\pi}}\int_0^{\infty} 2ye^{-y}\frac{\sqrt{2}}{2\sqrt{y}}dy = \frac{2}{\sqrt{\pi}}\int_0^{\infty} y^{\frac{3}{2}-1}e^{-y}dy = \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) \\
&= \frac{2}{\sqrt{\pi}} \times \frac{1}{2}\sqrt{\pi} = 1.
\end{aligned}$$

$$\begin{aligned}
E(X^4) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^4 e^{-x^2/2}dx = \sqrt{\frac{2}{\pi}}\int_0^{\infty} x^4 e^{-x^2/2}dx \\
\overset{y=\frac{x^2}{2}}{\underset{x=\sqrt{2y}}{=\!=\!=}}\ & \sqrt{\frac{2}{\pi}}\int_0^{\infty} 4y^2e^{-y}\frac{1}{\sqrt{2y}}dy = \frac{4}{\sqrt{\pi}}\int_0^{\infty} y^{\frac{5}{2}-1}e^{-y}dy = \frac{4}{\sqrt{\pi}}\Gamma\left(\frac{5}{2}\right) \\
&= \frac{4}{\sqrt{\pi}} \times \frac{3}{2} \times \frac{1}{2}\sqrt{\pi} = 3.
\end{aligned}$$

$$\operatorname{Var}(X^2) = E(X^4) - [E(X^2)]^2 = 2.$$

The remaining part is left as an exercise.

$X \sim N(0,1)$, $Y = X^2$,

$$Cov(X,Y) = Cov(X,X^2) = EX^3 - EXEX^2$$

Example 1.5.3. *Let* $p(x) = (1/2)^x, x = 1, 2, 3, \ldots$, *zero elsewhere, be the pmf of the r.v.* X. *Find the mgf, the mean and the variance of* X. *Let* $Y \sim \mathsf{Poisson}(\lambda)$. *Find the mgf, the mean and the variance of* Y.

Solution. Recall that $\sum_{i=1}^{\infty} a^i = \frac{a}{1-a}, \quad$ for $|a| < 1$. We have

$$M_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} \left(\frac{1}{2}\right)^x = \sum_{x=1}^{\infty} \left(\frac{e^t}{2}\right)^x = \frac{\frac{1}{2}e^t}{1 - \frac{1}{2}e^t} = \frac{e^t}{2 - e^t},$$

for $\frac{1}{2}e^t < 1$, or equivalently, $t < \log 2$. We have

$$\begin{aligned} M_X'(t) &= \frac{e^t(2 - e^t) + e^{2t}}{(2 - e^t)^2} = \frac{2e^t}{(2 - e^t)^2}, \\ M_X''(t) &= \frac{2e^t(2 - e^t)^2 + 2(e^t - 2)e^t \cdot 2e^t}{(2 - e^t)^4} = \frac{4e^t + 2e^{2t}}{(2 - e^t)^3}. \end{aligned}$$

Therefore $E(X) = M_X'(0) = 2, \quad E(X^2) = M_X''(0) = 6, \quad$ and $\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 2$. The remaining part is left as an exercise.

Remark 1.5.4. *Here, and in all the lecture notes of this subject, "$\log x$" is defined as the natural logarithm $\log_e x$, with the base $e = 2.71828182845904523536 0\ldots$*

Exercise 1.1

1. (1.6.9) Let X have a pmf $p(x) = 1/3$, $x = -1, 0, 1,$. Find the pmf of $Y = X^2$.
2. (1.7.6) For each of the following pdf's of X, find $P(|X| < 1)$ and $P(X^2 < 9)$.
   a. $f(x) = x^2/18$, $-3 < x < 3$, zero elsewhere.
   b. $f(x) = (x + 2)/18$, $-2 < x < 4$, zero elsewhere.
3. (1.7.7) Let $f(x) = 1/x^2$, $1 < x < \infty$, zero elsewhere, be the pdf of X. If $C_1 = \{x : 1 < x < 2\}$ and $C_2 = \{x : 4 < x < 5\}$, find $P(C_1 \cup C_2)$ and $P(C_1 \cap C_2)$.
4. (1.7.14) Let X have the pdf $f(x) = 2x$, $0 < x < 1$, zero elsewhere. Compute the probability that X is at least $\frac{3}{4}$ given the condition that X is at least $\frac{1}{2}$.
5. (1.7.20) Let X have the pdf $f(x) = x^2/9$, $0 < x < 3$, zero elsewhere. Find the pdf of $Y = X^3$.
6. (1.7.22) Let X have the uniform pdf $f_X(x) = \frac{1}{\pi}$, for $-\frac{\pi}{2} < x < \frac{\pi}{2}$. Find the pdf of $Y = \tan X$. This is the pdf of a **Cauchy distribution**.
7. (1.7.23) Let X have the pdf $f(x) = 4x^3$, $0 < x < 1$, zero elsewhere. Find the cdf and the pdf of $Y = -\log(X^4)$.
8. (1.7.24) Let $f(x) = \frac{1}{3}$, $-1 < x < 2$, zero elsewhere, be the pdf of X. Find the cdf and the pdf of $Y = X^2$.

1.6 Multivariate distributions

(1) Random vector: $\mathbf{X} = (X_1, X_2, \ldots, X_m)^T$. We here consider $m = 2$ only.

(2) Distributions

Joint cdf: $F_{X_1,X_2}(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2)$

Discrete case: Joint pmf of X_1 and X_2: $p_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$

Continuous case: Joint pdf $f_{X_1,X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1,X_2}(x_1,x_2)}{\partial x_1 \partial x_2}$

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(y_1, y_2) dy_1 dy_2$$

Marginals:

$F_{X_1}(x_1) = F_{X_1,X_2}(x_1, \infty)$

$p_{X_1}(x_1) = \sum_{x_2} p_{X_1,X_2}(x_1, x_2)$

$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) dx_2$

Transformations:

$$\left.\begin{array}{l} Y_1 = u_1(X_1, X_2) \\ Y_2 = u_2(X_1, X_2) \end{array}\right\} \overset{\text{one-to-one}}{\Longleftrightarrow} \left\{\begin{array}{l} X_1 = w_1(Y_1, Y_2) \\ X_2 = w_2(Y_1, Y_2) \end{array}\right.$$

Jacobian of the transformation

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix}$$

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(w_1(y_1, y_2), w_2(y_1, y_2))|J|$$

(3) Expectation

Let $Y = g(X_1, X_2)$

$$E(X) = \begin{cases} \displaystyle\sum_{x_1}\sum_{x_2} g(x_1, x_2)p(x_1, x_2) & \text{(discrete case)} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2)f(x_1, x_2)dx_1dx_2 & \text{(continuous case)} \end{cases}$$

Mean of $\mathbf{X}$ is $\mu = E(\mathbf{X}) = (E(X_1), E(X_2))^T$

$\mathrm{Cov}(X_1, X_2) = E[\{X_1 - E(X_1)\}\{X_2 - E(X_2)\}] = E(X_1X_2) - E(X_1)E(X_2)$

Correlation coefficient of X_1 and X_2

$$\rho = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_1\sigma_2}$$

(5) Conditional distributions

Conditional pmf: $p_{X_1,X_2|X_1}(x_1, x_2|x_1) = \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)}$

Conditional pdf: $f_{X_1,X_2|X_1}(x_1, x_2|x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}$

(6) Independence

X_1 and X_2 are said to be mutually independent

$$\Longleftrightarrow \begin{cases} p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2) & \text{(discrete case)} \\ f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) & \text{(continuous case)} \end{cases}$$

Question: Independent $\Rightarrow \rho = 0$? $\rho = 0 \Rightarrow$ Independent?

Example 1.6.1. *Let X_1 and X_2 have the joint pmf $p(x_1,x_2) = x_1x_2/36$, $x_1 = 1,2,3$, and $x_2 = 1,2,3$, zero elsewhere. Find first the joint pmf of $Y_1 = X_1X_2$ and $Y_2 = X_2$, and then find the marginal pmf of Y_1.*

Solution.

x_2 \ x_1	1	2	3
1	$(Y_1,Y_2)=(1,1)$ prob $= \frac{1}{36}$	$(Y_1,Y_2)=(2,1)$ prob $= \frac{2}{36}$	$(Y_1,Y_2)=(3,1)$ prob $= \frac{3}{36}$
2	$(Y_1,Y_2)=(2,2)$ prob $= \frac{2}{36}$	$(Y_1,Y_2)=(4,2)$ prob $= \frac{4}{36}$	$(Y_1,Y_2)=(6,2)$ prob $= \frac{6}{36}$
3	$(Y_1,Y_2)=(3,3)$ prob $= \frac{3}{36}$	$(Y_1,Y_2)=(6,3)$ prob $= \frac{6}{36}$	$(Y_1,Y_2)=(9,3)$ prob $= \frac{9}{36}$

The joint pmf of (Y_1, Y_2) is given by the following table:

Y_2 \ Y_1	1	2	3	4	6	9
1	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	0	0	0
2	0	$\frac{2}{36}$	0	$\frac{4}{36}$	$\frac{6}{36}$	0
3	0	0	$\frac{3}{36}$	0	$\frac{6}{36}$	$\frac{9}{36}$

The remaining part is left as an exercise.

Example 1.6.2. *Let* $f(x,y) = e^{-x-y}, 0 < x < \infty, 0 < y < \infty$, *zero elsewhere, be the pdf of* X *and* Y. *Let* $Z = X+Y$ *and* $W = X-Y$. *Compute the pdf of* Z *and* W.

Solution. We have $Z \geq 0$, so for any $t < 0$, $f_Z(t) = 0$. For $t \geq 0$, we have

$$
\begin{aligned}
F_Z(t) = P(Z \leq t) = \iint\limits_{\substack{x+y<t \\ x,y>0}} e^{-x-y} dx dy &= \int_0^t e^{-x} dx \int_0^{t-x} e^{-y} dy = \int_0^t e^{-x}(1 - e^{x-t}) dx \\
&= \int_0^t e^{-x} dx - \int_0^t e^{-t} dx = 1 - e^{-t} - te^{-t}.
\end{aligned}
$$

So $f_Z(t) = \frac{d}{dt} F_Z(t) = te^{-t}$. The remaining part is left as an exercise.

[this page left blank for the notes of Example 1.6.2]

Exercise 1.2

1. (2.1.1) Let $f(x_1, x_2) = 4x_1x_2$, $0 < x_1 < 1$, $0 < x_2 < 1$, zero elsewhere, be the pdf of X_1 and X_2. Find $P(0 < X_1 < 1/2, 1/4 < X_2 < 1)$, $P(X_1 = X_2)$, $P(X_1 < X_2)$, and $P(X_1 \leq X_2)$.

2. (2.1.6) Let $f(x, y) = e^{-x-y}, 0 < x < \infty, 0 < y < \infty$, zero elsewhere, be the pdf of X and Y. Then if $Z = X + Y$, compute $P(Z \leq 0), P(Z \leq 6)$, and more generally, $P(Z \leq z)$, for $0 < z < \infty$, what is the pdf of Z?

3. (2.1.7) Let X and Y have the pdf $f(x, y) = 1$, $0 < x < 1$, $0 < y < 1$, zero elsewhere. Find the cdf and pdf of the product $Z = XY$.

4. (2.1.16) Let X and Y have the joint pdf $f(x, y) = 6(1 - x - y)$, $x + y < 1$, $0 < x < y$, zero elsewhere. Compute $P(2X + 3Y < 1)$ and $E(XY + 2X^2)$.

5. (2.2.1) If $p(x_1, x_2) = \left(\frac{2}{3}\right)^{x_1+x_2} \left(\frac{1}{3}\right)^{2-x_1-x_2}$, $(x_1, x_2) = (0, 0), (0, 1), (1, 0), (1, 1)$, zero elsewhere, is the joint pmf of X_1 and X_2, find the joint pmf of $Y_1 = X_1 - X_2$ and $Y_2 = X_1 + X_2$.

6. (2.2.2) Let X_1 and X_2 have the joint pmf $p(x_1, x_2) = x_1x_2/36$, $x_1 = 1, 2, 3$, and $x_2 = 1, 2, 3$, zero elsewhere. Find first the pmf of $Y_1 = X_1X_2$ and $Y_2 = X_2$, and then find the marginal pmf of Y_1.

7. (2.2.3) Let X_1 and X_2 have the joint pdf $h(x_1, x_2) = 2\exp(-x_1 - x_2), 0 < x_1 < x_2 < \infty$, zero elsewhere. Find the joint pdf of $Y_1 = 2X_1$ and $Y_2 = X_2 - X_1$.

8. (2.2.4)* Let X_1 and X_2 have the joint pdf $h(x_1, x_2) = 8x_1x_2$, $0 < x_1 < x_2 < 1$, zero elsewhere. Find the joint pdf of $Y_1 = X_1/X_2$ and $Y_2 = X_2$.

9. (2.2.6) Suppose X_1 and X_2 have the joint pdf $f_{X_1,X_2}(x_1, x_2) = e^{-(x_1+x_2)}$, $0 < x_1, x_2 < \infty$, zero elsewhere. Find the pdf and mgf of $Y = X_1 + X_2$.

10. (2.3.8) Let X and Y have joint pdf $f(x, y) = 2\exp(-x - y), 0 < x < y < \infty$, zero elsewhere. Find the conditional mean $E(Y|x)$ of Y, given $X = x$.

11. Let $X, Y \sim \mathsf{N}(0, 1)$ be independent. Find the pdf of X^2. Find the pdf of X^2+Y^2.

12. Let $X, Y \sim \mathsf{N}(0, 1)$ be independent. Find the pdf of $2X$. Find the pdf of $X + Y$.

1.7 Summary of common discrete distributions

This list, as well as the list in the next section, serves as a reference, and the related part will be printed on the test and exam paper.

Bernoulli distribution; *Binomial* distribution; *Geometric* distribution; *Poisson* distribution

Bernoulli(p): $0 < p < 1$,

$$p(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$
$$\mu = p, \quad \sigma^2 = p(1-p), \quad M(t) = (1-p) + pe^t, \quad -\infty < t < \infty.$$

Binomial(n, p): $0 < p < 1$, and $n = 1, 2, \ldots$,

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$
$$\mu = np, \quad \sigma^2 = np(1-p), \quad M(t) = ((1-p) + pe^t)^n, \quad -\infty < t < \infty.$$

Geometric(p): $0 < p < 1$,

$$p(x) = p(1-p)^x, \quad x = 0, 1, 2, \ldots.$$
$$\mu = \frac{p}{1-p}, \quad \sigma^2 = \frac{1-p}{p^2}, \quad M(t) = p(1-(1-p)e^t)^{-1}, \quad t < -\log(1-p).$$

Hypergeometric(N, D, n): $n = 1, 2, \ldots, \min\{N, D\}$

$$p(x) = \frac{\binom{N-D}{n-x}\binom{D}{x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, n$$
$$\mu = n\frac{D}{N}, \quad \sigma^2 = n \times \frac{D}{N} \times \frac{N-D}{N} \times \frac{N-n}{N-1}.$$

Negative Binomial, NB(r, p): $0 < p < 1$, and $r = 1, 2, \ldots$.

$$p(x) = \binom{x+r-1}{r-1} p^r(1-p)^x, \quad x = 0, 1, 2, \ldots.$$
$$\mu = \frac{rp}{1-p}, \quad \sigma^2 = \frac{r(1-p)}{p^2}, \quad M(t) = p^r(1-(1-p)e^t)^{-r}, \quad t < -\log(1-p).$$

Poisson(λ): $\lambda > 0$

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots.$$
$$\mu = \sigma^2 = \lambda, \quad M(t) = \exp\{\lambda(e^t - 1)\}, \quad -\infty < t < \infty.$$

1.8 Summary of common continuous distributions

Uniform distribution; *Normal* distribution; *Exponential* distribution; *Gamma* distribution; *Chi-square* distribution; *Beta* distribution; t-distribution; F-distribution

Beta distribution, $\mathsf{Beta}(\alpha, \beta)$: $\alpha > 0$, and $\beta > 0$,

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1.$$

$$\mu = \frac{\alpha}{\alpha+\beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

$$M(t) = 1 + \sum_{i=1}^{\infty} \left(\prod_{j=0}^{i-1} \frac{\alpha+j}{\alpha+\beta+j} \right) \frac{t^i}{i!}, \quad -\infty < t < \infty.$$

Cauchy:

$$f(x) = \frac{1}{\pi(x^2+1)}, \quad -\infty < x < \infty.$$

Neither the mean nor the variance exists. The MGF does not exist.
Chi-squared distribution, $\chi^2(r)$: $r > 0$. The cases $r = 1, 2, \ldots$ are usually used.

$$f(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} x^{(r/2)-1} e^{-x/2}, \quad x > 0.$$

$$\mu = r, \quad \sigma^2 = 2r, \quad M(t) = (1-2t)^{-r/2}, \quad t < \frac{1}{2}.$$

$$X \sim \chi^2(r) \iff X \sim \mathsf{Gamma}\left(\frac{r}{2}, 2\right).$$

Exponential(λ): $\lambda > 0$

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

$$\mu = \frac{1}{\lambda}, \quad \sigma^2 = \frac{1}{\lambda^2}, \quad M(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}.$$

$$X \sim \mathsf{Exponential}(r) \iff X \sim \mathsf{Gamma}\left(1, \frac{1}{\lambda}\right).$$

F distribution, $\mathsf{F}(r_1, r_2)$: $r_1 > 0$ is called the numerator degrees of freedom, and $r_2 > 0$ is called the denominator degrees of freedom.

$$f(x) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{r_1/2}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)} \frac{x^{(r_1/2)-1}}{(1+r_1x/r_2)^{(r_1+r_2)/2}}, \quad x > 0.$$

$$\text{If } r_2 > 2, \mu = \frac{r_2}{r_2-2}; \quad \text{If } r_2 > 4, \sigma^2 = 2\left(\frac{r_2}{r_2-2}\right)^2 \frac{r_1+r_2-2}{r_1(r_2-4)}.$$

The MGF does not exist.

$\mathsf{Gamma}(\alpha, \beta)$: $\alpha > 0$ and $\beta > 0$.

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0.$$

$$\mu = \alpha\beta, \quad \sigma^2 = \alpha\beta^2, \quad M(t) = (1 - \beta t)^{-\alpha}, \quad t < \frac{1}{\beta}.$$

$$X_1, \ldots, X_m \sim \mathsf{Gamma}(\alpha, \beta) \text{ and independent} \Longrightarrow \sum_{i=1}^m X_i \sim \mathsf{Gamma}(m\alpha, \beta).$$

$\mathsf{Laplace}(\theta)$: $-\infty < \theta < \infty$.

$$f(x) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty.$$

$$\mu = \theta, \quad \sigma^2 = 2, \quad M(t) = e^{t\theta} \frac{1}{1 - t^2}, \quad -1 < t < 1$$

$\mathsf{Logistic}(\theta)$: $-\infty < \theta < \infty$.

$$f(x) = \frac{\exp(-(x - \theta))}{(1 + \exp(-(x - \theta)))^2}, \quad -\infty < x < \infty.$$

$$\mu = \theta, \quad \sigma^2 = \frac{\pi^2}{3}, \quad M(t) = e^{t\theta}\Gamma(1 - t)\Gamma(1 + t), \quad -1 < t < 1.$$

Normal, $\mathsf{N}(\mu, \sigma^2)$: $-\infty < \mu < \infty$, and $\sigma > 0$.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}, \quad -\infty < x < \infty.$$

$$\mu = \mu, \quad \sigma^2 = \sigma^2, \quad M(t) = \exp\left\{ \mu t + \frac{1}{2}\sigma^2 t^2 \right\}.$$

t-distribution, $t(r)$, $r > 0$.

$$f(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r \Gamma\left(\frac{r}{2}\right)}} \cdot \frac{1}{\left(1 + \frac{x^2}{r}\right)^{(r+1)/2}}, \quad -\infty < x < \infty.$$

$$\text{if } r > 1, \quad \mu = 0. \quad \text{If } r > 2, \quad \sigma^2 = \frac{r}{r - 2}.$$

The mgf does not exist.

The parameter r is called the degrees of freedom.

<u>Uniform(a, b)</u>, $-\infty < a < b < \infty$.

$$f(x) = \frac{1}{b-a}, \quad a < x < b.$$

$$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}, \quad M(t) = \frac{e^{bt} - e^{at}}{(b-a)t}, \quad -\infty < t < \infty.$$

1.9 Monty Hall problem

It became famous as a question from a reader's letter quoted in Marilyn vos Savant's "Ask Marilyn" column in Parade magazine in 1990:

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?