Random Variables and Their Distributions

Bo Li, Spring 2019

(I) Let $(\Omega, \mathcal{A}, P)$ be a probability space (i.e., $\Omega \neq \phi$ is a set, called the sample space and $\omega \in \Omega$ is called a sample point or elementary event; $\mathcal{A}$ is a σ-algebra, consisting of subsets of Ω and each member in $\mathcal{A}$ is called an event; and P is the probability measure).

Definition A (real-valued, scalar) random variable (RV) is a mapping $X : \Omega \to \mathbb{R}$ such that

$$\{X \le x\} := \{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{A} \quad \forall x \in \mathbb{R}$$

Equivalently, $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$ for any $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra of $\mathbb{R}$

Definition The expectation (or mean) of a RV $X : \Omega \to \mathbb{R}$ is $\langle X \rangle = \mathbb{E}(X) = \mathbb{E}X = \int_\Omega X \, dP$, if the integral exists.

The variance of $X : \Omega \to \mathbb{R}$ is

$$Var(X) = \mathbb{E}((X - \mathbb{E}X)^2) \quad \text{if } \mathbb{E}|X|^2 < \infty.$$

Proposition (Prove all these!)

① $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y \quad a, b \in \mathbb{R}, X, Y : \text{RVs}$

② $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ Important!

③ $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$, if X, Y are independ.

Also, $\mathbb{E}1 = 1$. $\mathbb{E}X \ge 0$ if $X \ge 0$ on Ω.

Put cov(X, Y) here.

Definition X: a RV. The <u>kth moment</u> of X is $m_k = E(X^k)$ (if exists). The kth central ~~central~~ moment is $\mu_k = E((X - E(X))^k)$.

Definition The <u>cumulative distribution function</u> (CDF) of a RV $X: \Omega \to \mathbb{R}$ is $F_X: \mathbb{R} \to [0,1]$ (or $F: \mathbb{R} \to [0,1]$):

Or: just distribution

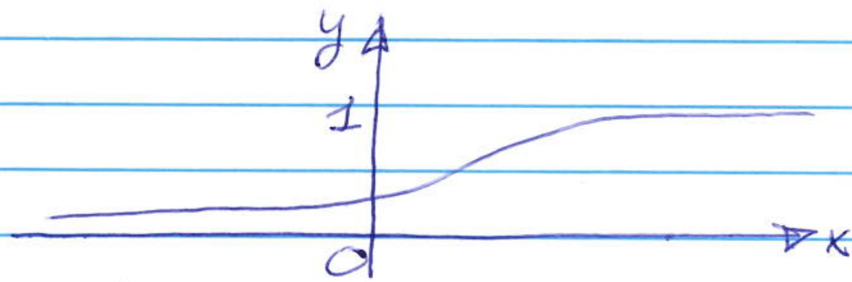
$$F_X(x) = P(X \le x) = P(\{\omega \in \Omega : X(\omega) \le x\}) \quad \forall x \in \mathbb{R}$$

Proposition ① F_X is nondecreasing

② $F_X(-\infty) = 0$ and $F_X(\infty) = 1$

③ F_X is right-continuous.

If a function $\mathbb{R} \to [0,1]$ satisfies ①, ② and ③ this it is a CDF for a RV.



Proposition ① $P(X > x) = 1 - F_X(x) \quad \forall x \in \mathbb{R}$

Prove all these!

② $P(x < X \le y) = F(y) - F(x) \quad \forall x, y \in \mathbb{R},\ x < y$.

Note: the subtle point here.

③ $P(X = x) = F_X(x) - \underbrace{F_X(x-)}_{\lim_{z \to x-} F_X(z)} \quad \forall x \in \mathbb{R}$

(II) **Definition** A RV $X: \Omega \to \mathbb{R}$ is <u>discrete</u> if the range of X, $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$, is finite or countably infinite.

If X is a discrete RV, then there exist finitely many or countably infinitely many distinct $x_1, x_2, \cdots \in \mathbb{R}$ such that

$$X(\Omega) = \{x_1, x_2, \cdots\}.$$

For each i, let $\Omega_i = \{\omega \in \Omega : X(\omega) = x_i\} = X^{-1}(\{x_i\})$

Then $\Omega = \bigcup_{i \geq 1} \Omega_i$, $\Omega_i \cap \Omega_j = \phi \quad (i \neq j)$

$$X = \sum_i x_i \mathbb{1}_{\Omega_i}$$

Here and below, for any $A \in \mathcal{A}$, $\mathbb{1}_A : \Omega \to \mathbb{R}$ is the <u>indicator function</u> of A

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Or, just the density

<u>Definition</u> The <u>probability mass function</u> (PMF), or the <u>discrete probability ~~distribution~~ density function</u> (DPDF), or simply <u>discrete density function</u> (DDF), of a discrete RV $X : \Omega \to \mathbb{R}$ is a function

$$f_X : \mathbb{R} \to [0, 1] \quad \text{defined by}$$

$$f_X(x) = P(X = x).$$

<u>~~Property~~ Proposition</u>

Try to prove all these except ④ on the next page.

① If $X(\Omega) = \{x_1, x_2, \cdots\}$ then

$$F_X(x) = \sum_{i : x_i \leq x} f_X(x_i)$$

② In general $f_X(x) = F_X(x) - F_X(x-)$

~~Property~~ Proposition. The PMF $f_X : \mathbb{R} \to [0,1]$ of a discrete RV X with $X(\Omega) = \{x_1, x_2, \cdots\}$ satisfies

① $f_X(x) \neq 0 \iff x \in \{x_1, x_2 \cdots\}$.

② $\sum_i f_X(x_i) = 1$.

Moreover, ①, ② characterize a PMF.

Calculations of the mean and variance of a discrete RV X can be done through is PMF f_X.

Suppose $X(\Omega) = \{x_1, x_2, \cdots\}$. Then

① $\mathbb{E}(X) = \sum_x x f_X(x) = \sum_i x_i f_X(x_i)$

② $\mathbb{E}(g(X)) = \sum_i g(x_i) f_X(x_i)$

Prove this using $X = \sum_i x_i \mathbb{1}_{\Omega_i}$ where $\Omega_i = X^{-1}(\{x_i\})$

Here, we assume the sums, if they are infinite, converge absolutely.

(III) **Definition** A RV $X : \Omega \to \mathbb{R}$ is continuous (i.e., X is a continuous RV) if there exists $f_X : \mathbb{R} \to [0, \infty)$ s.t.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad \forall x \in \mathbb{R}.$$

We call f_X is probability density function (PDF) of the RV. Or, just the density

~~Property~~ Proposition ① $f_X(x) = F_X'(x)$ at x at which f_X is cont.

② $f_X(x) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

③ $P(a \le X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(t)\,dt$ $\quad \forall a, b \in \mathbb{R}$, $a < b$

④ $P(X \in B) = \int_B f_X(x)\,dx$ for any Borel set $B \in \mathcal{B}$

where $\mathcal{B}$ is the Borel σ-algebra of $\mathbb{R}$

⑤ $P(X = x) = 0 \quad \forall x \in \mathbb{R}$

Moreover, any $h : \mathbb{R} \to [0, \infty)$, with $\int_{-\infty}^{\infty} h(x)\,dx = 1$, can serve as a PDF for some continuous RV.

For a continuous RV, X, we can calculate its expectation and variance using the PDF f_X. We have

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

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \quad \text{if } g \text{ is a continuous function}$$

Here, we assume the integrals converge absolutely.

(IV) Examples of Distributions of Discrete RVs

1. Bernoulli variables and Bernoulli distribution.

A discrete RV X is called a Bernoulli variable if $X(\Omega) = \{0, 1\}$.

Let $p = P(X = 1)$ and $q = P(X = 0) = 1 - p$. $(0 < p < 1)$

Then the PMF f_X is given by

$$f_X(x) = 0 \text{ if } x \neq 0, 1, \quad f_X(0) = q, \quad f_X(1) = p.$$

This is the Bernoulli distribution : Bernoulli(p)

Note: The CDF is $F_X(x) = \sum_{k:\, x_k \le x} f_X(x_k) = \begin{cases} 0 & \text{if } x<0, \\ 1-p & \text{if } 0 \le x < 1, \\ 1 & \text{if } 1 \le x < \infty. \end{cases}$ 6

We have

$$\mathbb{E}(X) = p. \quad Var(X) = pq.$$

If $A \in \mathcal{A}$ then the indicator variable $\mathbb{I}_A$ is a Bernoulli variable

$$\mathbb{E}(\mathbb{I}_A) = P(\mathbb{I}_A = 1) = P(A)$$

$$Var(\mathbb{I}_A) = P(A)\, P(A^c)$$

We write $X \sim$ Bernoulli(p) to indicate that X is a Bernoulli variable and $P(X=1) = p$.

The set of all natural numbers $1, 2, 3, \dots$

2. Binomial distribution $B(n,p)$ $\quad (n \in \mathbb{N},\ 0<p<1)$

A discrete RV $X : \Omega \to \{0, 1, 2, \dots, n\}$ is said to have the binomial distribution with parameters n, p, written $X \sim B(n,p)$ if the PMF $f_X : \mathbb{R} \to [0,1]$ is given by

$$f_X(x) = 0 \quad \text{if } x \notin \{0, 1, \dots, n\}$$

$$f_X(k) = \binom{n}{k} p^k q^{n-k} \quad \text{if } k \in \{0, 1, \dots, n\}$$

where $q = 1-p$

It is clear that $f_X \ge 0$ and

$$\sum_{k=0}^{n} f_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1$$

The CDF is $F_X(x) = \cancel{\begin{cases} 0 & \text{if } x<0 \\ \sum_{0 \le k \le x} \binom{n}{k} p^k q^{n-k} & \text{if } 0 \le x < \infty \end{cases}}$ $0 \le x \le n$

If $X \sim B(n,p)$, we say X is a Binomial variable.

We have for $X \sim B(n,p)$.

$$\mathbb{E}(X) = np \quad \text{and} \quad Var(X) = np(1-p) = npq$$

The CDF is $F_X(x) = \begin{cases} 0 & \text{if } x<0 \\ \sum_{0 \le k \le x} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } 0 \le x \le n \\ 1 & \text{if } x \ge n \end{cases}$

3. <u>Poisson variable and Poisson distribution $P(\lambda)$ with parameter $\lambda>0$.</u>

A discrete RV X is a Poisson variable if it has a Poisson distribution with some parameter $\lambda>0$, written $X\sim \text{Poisson}(\lambda)$ or $P(\lambda)$,

$$X:\Omega\longrightarrow\{0,1,2,\cdots\}$$

$$f_X(x)=0 \quad \text{if } x\notin\{0,1,2,\cdots\}$$

$$f_X(k)=\frac{\lambda^k}{k!}e^{-\lambda} \quad \text{if } k\in\{0,1,2,\cdots\}.$$

Note That

$$\sum_{k=0}^{\infty} f_X(k)=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1.$$

We have

The CDF is
$$F_X(x)=\begin{cases}0 & \text{if } x<0\\ \sum\limits_{0\le n\le x}\frac{\lambda^n e^{-\lambda}}{n!} & \text{if } 0\le x<\infty\end{cases}$$

$$\mathbb{E}(X)=\sum_{k=0}^{\infty}k\,f_X(k)=\sum_{k=1}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}$$

$$=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

$\mathbb{E}(X)=\lambda$ and $Var(X)=\lambda$

Prove this.

4. <u>Geometric variable and geometric distribution with parameter $p\in(0,1)$.</u>

We say a discrete RV $X:\Omega\to\{1,2,\cdots\}$ is a geometric variable, with the geometric distribution with parameter $p\in(0,1)$ written <u>$X\sim \text{Geometric}(p)$</u> if the PMF of X is

The CDF is
$$F_X(x)=\begin{cases}0 & \text{if } x<0\\ \sum\limits_{0\le n\le x} p(1-p)^n & \text{if } x\ge 0\end{cases}$$

$$f_X(x)=0 \quad \text{if } x\notin\{1,2,\cdots\}$$

$$f_X(k)=p(1-p)^{k-1},\quad k=1,2,\cdots$$

We have $\mathbb{E}(X)=\frac{1}{p}$ and $Var(X)=\frac{1-p}{p^2}$.

Prove these!

5. The discrete uniform distribution over $\{1,2,\cdots,n\}$.

A discrete RV $X:\Omega \to \{1,2,\cdots,n\}$ with such a distribution has the PMF

$$f_X(x) = \begin{cases} 0 & \text{if } x \notin \{1,2,\cdots,n\} \\ \frac{1}{n} & \text{if } x \in \{1,2,\cdots,n\}. \end{cases}$$

The CDF is

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{[x]}{n} & \text{if } 0 \le x \le n \\ 1 & \text{if } x > n \end{cases}$$

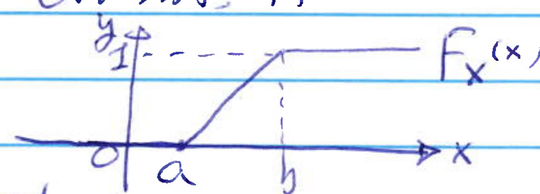
where $[x]$ is the largest integer $\le x$.

(V) Examples of Distributions of Continuous RVs

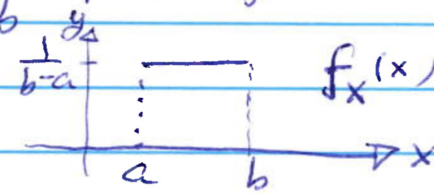
1. The uniform distribution on $[a,b]$ $(a,b \in \mathbb{R},\ a<b)$

A continuous RV $X:\Omega \to \mathbb{R}$ with such a distribution, written $X \sim U[a,b]$, if

$$F_X(x) = \begin{cases} 0 & \text{if } x \le a \\ \frac{x-a}{b-a} & \text{if } a < x \le b \\ 1 & \text{if } x > b \end{cases}$$



$$f_X(x) = \begin{cases} 0 & \text{if } x \le a \text{ or } x > b \\ \frac{1}{b-a} & \text{if } a < x < b \end{cases}$$



check these!

$$\begin{cases} \mathbb{E}(X) = \frac{1}{2}(a+b). \\ Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{(b-a)^2}{12} \end{cases}$$

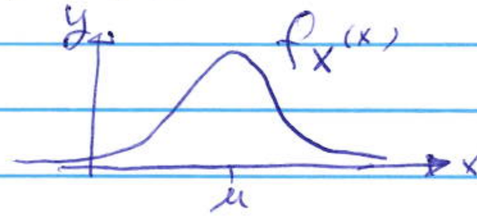
2. The normal distribution (or Gaussian distribution) $\mathcal{N}(\mu, \sigma^2)$ $(\mu \in \mathbb{R}, \sigma > 0)$

A continuous RV X that has such a distribution, means that

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

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt. \qquad \forall x \in \mathbb{R}.$$

Verify that $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.



Verify these! $\mathbb{E}(X) = \mu$

$\mathrm{Var}(X) = \sigma^2$

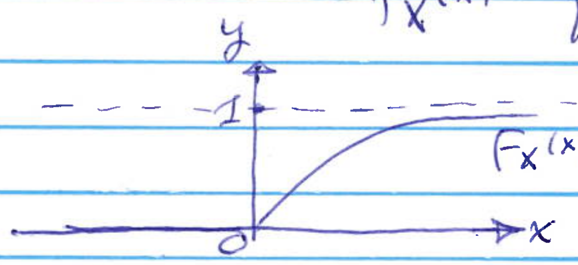
The standard normal distribution is $\mathcal{N}(0,1)$

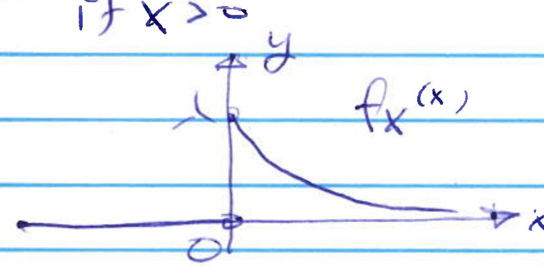
3. The exponential distribution with parameter $\lambda > 0$ denoted $\mathrm{Exp}(\lambda)$.

$X \sim \mathrm{Exp}(\lambda)$: X is a continuous RV with the exponential distribution $\mathrm{Exp}(\lambda)$

It means: $F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x > 0 \end{cases}$$





$\mathbb{E}(X) = \frac{1}{\lambda}$. $\left[\text{since } \mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^{\infty} x \lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}\right]$.

$\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

4. The Gamma Distribution $\Gamma(\lambda, t)$ $(\lambda, t > 0)$

A RV X has such a distribution, $X \sim \Gamma(\lambda, t)$, if

$$f_X(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} \quad \text{if } x \geq 0$$
$$f_X(x) = 0 \quad \text{if } x < 0.$$

where $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\,dx$ is the Gamma function.

$t = 1$: This is $Exp(\lambda)$.

5. The Beta Distribution $B(a,b)$ $(a, b > 0)$

A RV X has such a distribution, $X \sim B(a,b)$, if

$$f_X(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \quad 0 \leq x \leq 1.$$

$$f_X(x) = 0 \text{ if } x < 0 \text{ or } x > 1.$$

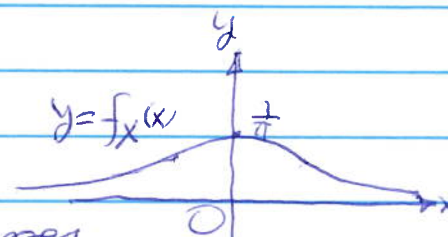
where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$

is the Beta function. $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

6. The Weibull distribution $W(\alpha, \beta)$ $(\alpha, \beta > 0)$

$X \sim W(\alpha, \beta)$:
$$f_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} & \text{if } x \geq 0. \end{cases}$$

7. The Cauchy Distribution.



$$f_X(x) = \frac{1}{\pi(1+x^2)} \quad (x \in \mathbb{R})$$

No $E(X)$ — $\int_{-\infty}^\infty x f_X(x)\,dx$ diverges.

(VI) Joint Distributions for Multiple RVs

Definition Let $X_1, X_2, \ldots, X_n$ be n real-valued scalar RVs. The joint cumulative distribution function (joint CDF), or simply, the joint distribution of these RVs is $F = F_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$, defined by

$$F(x_1, \ldots, x_n) = \mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) \quad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

Here, $(X_1 \le x_1, \ldots, X_n \le x_n)$

$$= \{\omega \in \Omega : X_1(\omega) \le x_1, X_2(\omega) \le x_2, \ldots, X_n(\omega) \le x_n\}.$$

$$= \bigcap_{k=1}^{n} \{\omega : X_k(\omega) \le x_k\}.$$

Definition If $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ are discrete RVs, then the joint probability mass function (joint PMF), or the joint density/discrete function of these RVs is $f = f_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$, given by

$$f(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

$$= \mathbb{P}\left(\bigcap_{j=1}^{n} \{X_j = x_j\}\right) \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Calculate the mean/expectation of a function of multiple discrete RVs.

Proposition ~~Lemma~~ If $X_1, \ldots, X_n$ are real-valued, scalar, discrete RVs then for any continuous function $g : \mathbb{R}^n \to \mathbb{R}$, $g(X_1, \ldots, X_n)$ is also a discrete RV. Moreover,

$$\mathbb{E}\left(g(X_1, \ldots, X_n)\right) = \cancel{\sum_{i=1}} \sum_{x_1,\ldots,x_n} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n),$$

if the sum is finite, or countably infinite which converges absolutely.

Definition Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be continuous RVs with the joint distribution (i.e. the joint CDF) $F = F_{X_1, \cdots, X_n} : \mathbb{R}^n \to [0,1]$. Suppose

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \int_{-\infty}^{x_n}\int_{-\infty}^{x_{n-1}}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f_{X_1,\cdots,X_n}(u_1,\cdots,u_n)\,du_1\,du_2\cdots du_n$$

for some integrable function

$$f = f_{X_1,\cdots,X_n} : \mathbb{R}^n \to [0,\infty)$$

then we call $f_{X_1,\cdots,X_n}$ the joint PDF for $X_1, \cdots, X_n$

Note that ① $f_{X_1,\cdots,X_n} \geq 0$ and ② $\int_{\mathbb{R}^n} f(x)\,dx = 1$

Proposition

③ $f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \dfrac{\partial^n F_{X_1,\cdots,X_n}(x_1,\cdots,x_n)}{\partial x_1 \cdots \partial x_n}$

if the partial derivative exists.

④ $P(\vec{X} \in B) = \int_B f_{X_1,\cdots,X_n}(x)\,dx$

$\vec{X} = (X_1, \cdots, X_n)$, $B \subseteq \mathbb{R}^n$: a Borel set.

In particular,

$$P(a_1 \leq X_1 \leq b_1, \cdots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1}\cdots\int_{a_n}^{b_n} f_{X_1\cdots X_n}(x_1,\cdots,x_n)\,dx_1\cdots dx_n.$$

<u>Proposition</u> If $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ are continuous RVs with the joint density $f_{X_1,\cdots,X_n} : \mathbb{R}^n \to [0,\infty)$. and $g : \mathbb{R}^n \to \mathbb{R}$ is a nice function (e.g., smooth, integrable on $\mathbb{R}^n$, etc.), then

$g(X_1, \cdots, X_n)$ is a ~~contin~~ RV and

$$\mathbb{E}(g(X_1,\cdots,X_n)) = \int_{\mathbb{R}^n} g(x_1,\cdots,x_n) f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)\,dx_1\cdots dx_n$$

$$= \int_{\mathbb{R}^n} g(x) f_{\underline{X}}(x)\,dx.$$

$(\ x = (x_1, \cdots, x_n)\ )$

<u>Marginal Distributions</u>

"<u>Definition</u>" Marginal distributions are just distributions for a single RV from a group of RVs.

Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be RVs. <u>The marginal distribution function</u> for the ith RV X_i is given by

$$F_{X_i}(x_i) = P(X_i \le x_i) = F_{X_1,\cdots,X_n}(+\infty, \cdots, +\infty, x_i, +\infty, \cdots, +\infty)$$

$$= \lim_{\substack{x_j \to +\infty \\ j=1,\cdots,n,\ j \ne i}} F_{X_1,\cdots,X_n}(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n).$$

where $F_{X_1,\cdots,X_n} : \mathbb{R}^n \to [0,1]$ is the joint distribution function for $X_1, \cdots, X_n$.

$n=2$, $X = X_1$, $Y = X_2$

$F_X(x) = F_{X,Y}(x, \infty)$, $F_Y(y) = F_{X,Y}(\infty, y)$ $\forall x, y \in \mathbb{R}$

If X, Y are discrete RVs, then the marginal prob. mass functions f_X, f_Y are given by

$$f_X(x) = \sum_y f_{X,Y}(x,y),$$

$$f_Y(y) = \sum_x f_{X,Y}(x,y),$$

where $f_{X,Y}(x,y) = P(X=x, Y=y)$ is the joint mass function.

If X, Y are continuous RVs, then the marginal PDF are

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

where $f_{X,Y}$ is the joint PDF (the joint density) for X, Y.

(VII) Independence

Definition Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $A, B \in \mathcal{A}$ (two events) with $P(B) > 0$. The conditional probability of A given B (i.e. the probility A occurs given that B occurs) is

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

Lemma (1) If $P(B)>0$ ~~then~~ and $P(B^c)>0$ then

$$P(A)=P(A|B)P(B)+P(A|B^c)P(B^c)$$

(2) If $B_1,\dots,B_n$ is a partition of Ω (i.e., all $B_j \in \Omega$, $j=1,\dots,n$, $B_i \cap B_j = \emptyset \ \forall i \neq j$, $\bigcup_{i=1}^{n} B_i = \Omega$) and $P(B_j)>0$ ($j=1,2,\dots,n$), then

$$P(A)=\sum_{i=1}^{n} P(A|B_i)P(B_i) \quad \forall A \in \mathcal{A}.$$

Definition Two events $A, B \in \mathcal{A}$ are independent, if $P(A \cap B) = P(A)P(B)$.

A family of events $A_i \in \mathcal{A}$ ($i \in I$) is independent if

$$P\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} P(A_i)$$

for any finite subset J of I.

Definition Let $X_1, X_2, \dots, X_n : \Omega \to \mathbb{R}$ be RVs. They are independent if

$$P\Big(\bigcap_{i=1}^{n} \{X_i \in B_i\}\Big) = \prod_{i=1}^{n} P(\{X_i \in B_i\})$$

for any Borel sets $B_1, \dots, B_n$ of $\mathbb{R}$.

Theorem ~~Proposition~~ The RVs $X_1, \dots, X_n : \Omega \to \mathbb{R}$ are indep. $\iff$ (if and only if)

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad \forall x_1,\dots,x_n \in \mathbb{R}$$

$$\iff$$

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

[$f_{X_1,\dots,X_n}$, f_{X_i} can be discrete or continuous densities.]

Theorem If $X_1, \dots, X_n$ are independent RVs and $g: \mathbb{R}^k \to \mathbb{R}$ and $h: \mathbb{R}^{n-k} \to \mathbb{R}$ are two functions, where $1 \le k \le n-1$, then $g(X_1, \dots, X_k)$ and $h(X_{k+1}, \dots, X_n)$ are independent RVs.

Theorem Let $X_1, \dots, X_n$ be independent RVs. Then

(1) $\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i)$

provided that each $\mathbb{E}(|X_i|) < \infty$.

(2) $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \stackrel{!}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i)$

provided that each $\mathrm{Var}(X_i) < \infty$.

Note: If X, Y are discrete RVs,

$X(\Omega) = \{x_1, x_2, \dots\}$,

$Y(\Omega) = \{y_1, y_2, \dots\}$

Then $X = \sum_i x_i \mathbb{1}_{A_i}$ $\quad A_i = \{X = x_i\}$

$Y = \sum_j y_j \mathbb{1}_{B_j}$ $\quad B_j = \{Y = y_j\}$.

X, Y are independent $\Leftrightarrow$ A_i, B_j are independent for any i, j

Some skipped subjects.

Move this part on $\mathrm{cov}(X,Y)$ to right before ~~the~~ def. of CDF on page $\boxed{2}$.

(~~VIII~~)

(~~VIII~~) **Definition** The covariance of two RVs X and Y is

$$\mathrm{cov}(X,Y) = \mathbb{E}\big((X - E(X))(Y - E(Y))\big).$$

The correlation (coefficient) of X and Y is

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}.$$

We have

① $\mathrm{cov}(X, Y) = \mathbb{E}(XY) - E(X)\,E(Y)$

② $|\rho(X, Y)| \le 1$.

"=" holds $\Longleftrightarrow$ $P(Y = aX + b) = 1$ for some $a, b \in \mathbb{R}$.

Note ② follows from the Cauchy-Schwarz inequality:

$$[\mathbb{E}(XY)]^2 \le \mathbb{E}(X^2)\,E(Y^2).$$

Move this part to right before (VII) Independence, page $\boxed{14}$

The bivariate normal distribution ~~for a pair of RVs X and Y~~ $f : \mathbb{R}^2 \to [0,\infty)$

$$f(x, y) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}}\, e^{-\frac{1}{2} Q(x,y)}$$

$$Q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

where $\sigma_1, \sigma_2 > 0$, $0 < \rho < 1$, $\mu_1, \mu_2 \in \mathbb{R}$.

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

Prove these!

② X is $N(\mu_1, \sigma_1^2)$ and Y is $N(\mu_2, \sigma_2^2)$

③ The correlation between X and Y is ρ.

④ X, Y are independent $\Longleftrightarrow$ $\rho = 0$.

The standard bivariate normal distribution for X and Y is

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho x y + y^2)}$$

$[\mu_1 = \mu_2 = 0,\ \sigma_1 = \sigma_2 = 1]$

Here, $\mathrm{cov}(X, Y) = \rho$.

Definition Let $\mu = (\mu_1, \cdots, \mu_n) \in \mathbb{R}^n$, $V = (v_{ij})_{n \times n}$ be an $n \times n$, real, symmetric, nonsigular matrix. RVs $X_1, X_2, \cdots, X_n : \Omega \to \mathbb{R}$ has the <u>multivariate normal distribution</u> (or multinormal distribution), written $N(\mu, V)$, if their joint density function $f = f_{X_1, \cdots, X_n}$ is given by

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(V)}} e^{-\frac{1}{2}(x-\mu)V^{-1}(x-\mu)}$$

for all $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$.

Call $V = (v_{ij})$ the covariant matrix.

Theorem If $X = (X_1, \cdots, X_n)$ is $N(\mu, V)$, then

(1) $\mathbb{E}(X) = \mu$, i.e., $\mathbb{E}(X_i) = \mu_i$, $i = 1, 2, \cdots, n$;

(2) For $V = (v_{ij})$, we have

$$v_{ij} = \operatorname{cov}(X_i, X_j), \quad i, j = 1, 2, \cdots, n.$$

(VIII) Distributions of Functions of RVs and Sums of RVs

If $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ are RVs, then $X = (X_1, \cdots, X_n) : \Omega \to \mathbb{R}^n$ is a vector-valued RV. The distribution of this RV X is the joint distribution of $X_1, \cdots, X_n$ with the density

$$f_X(x) = f_{X_1, \cdots, X_n}(x) \quad \forall x = (x_1, \cdots, x_n) \in \mathbb{R}^n.$$

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a C^1-map ~~with the Jacobian~~ that is invertible from Range $T \to \mathbb{R}^n$. We denote by $J = J(y_1, \cdots, y_n) = J(y)$ the Jacobian of T^{-1}, the inverse of T.

$$T(x_1, x_2, \dots, x_n) = (y_1, \dots, y_n) \iff x_i = x_i(y_1, \dots, y_n) \quad i=1,\dots,n.$$

$$J(y) = \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right| = \det\left(\frac{\partial x_i}{\partial y_j} \right).$$

<u>Theorem</u> Under these assumptions the ~~dis~~ joint density function for $Y = (Y_1, \dots, Y_n) = T(X) = T(X_1, \dots, X_n)$ is

$$f_Y(y) = \begin{cases} f_X(T^{-1}(y))\,|J(y)| & \text{if } y \in \text{Range}(T) \\ 0 & \text{otherwise}. \end{cases}$$

<u>Theorem</u> Let $X, Y : \Omega \to \mathbb{R}$ be two RVs having the joint density $f_{X,Y}$.

(1) Let $Z = X + Y$. Then Z has the density
$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx \quad \forall z \in \mathbb{R}.$$

(2) If X, Y are independent, then
$$f_{X+Y} = f_X * f_Y \quad \text{(convolution)}.$$

Some examples of distributions of sums of RVs or functions of RVs.

(1) Let $X_1, \dots, X_n : \Omega \to \mathbb{R}$ be independent RVs. Assume X_i is Gamma(a_i, b) distributed, $i=1,\dots,n$, $a_i, b > 0$. Then $X_1 + \dots + X_n$ is Gamma$(\sum_{i=1}^{n} a_i, b)$ distributed.

(2) If a RV $Z : \Omega \to \mathbb{R}$ is $N(0,1)$ distributed, then Z^2 is Gamma$(\frac{1}{2}, \frac{1}{2})$ distributed.

(3) If $Z_1, \ldots, Z_n$ are i.i.d. $N(0,1)$ RVs, then $Z_1^2 + \cdots + Z_n^2$ is Gamma$(n/2, 1/2)$ distributed. (This is also called the chi-squared distribution with degree ~~k~~ n denoted χ_n^2.

(4) If $X_1, \ldots, X_n$ are i.i.d. with common distribution Bernoulli (p), then $X_1 + \cdots + X_n$ is $B(n, p)$ distributed.