Stat 240 Notes
By Marty Mukherjee

1. Foundations

Probability space: A set of objects (Ω, F, P):

Ω: Sample space

F: σ-algebra

P: Probability

1.1 σ-algebra and measure theory

<u>Definition 1.3: σ-algebra</u>

$p(\Omega)$ is the power set

$F \subseteq p(\Omega)$ is a σ-algebra on Ω if:

i) $\Omega \in F$

ii) $A \in F \implies A^c \in F$

iii) $A, B \in F \implies A \cup B \in F$
If (iii) holds for finitely many sets, then F is an algebra on Ω.

<u>Remark 1.4</u>

σ-algebras are closed with respect to countable intersections since by De Morgan's Law:

$$\cup_{i=1}^{\infty} A_i = (\cap_{i=1}^{\infty} A_i^c)^c$$

<u>Definition 1.7: Borel σ-algebra</u>

$B(\Omega)$ is a Borel σ-algebra on Ω and it's elements are Borel sets

$B(\Omega) = \sigma(\{A : A \subseteq \Omega, A \text{ is open}\})$

$B(\mathbb{R}^d) = \sigma(\{(\underline{a}, \underline{b}] : \underline{a} \leq \underline{b}\})$

<u>Definition 1.9: Measures</u>

Let $F = \sigma(\Omega)$. (Ω, F) is a measurable space. And sets in F are measurable sets.

A measure $\mu(F)$ is a μ if:

i) $\mu : F \to [0, \infty]$

ii) $\mu(\phi) = 0$

iii) σ-additivity

(Ω, F, μ) is called a measure space.

1.2 Probability Measures

Definition 1.9: Probability measure

Let (Ω, F) be a measurable space. A probability measure P on F is such that:

i) $P : F \to [0, 1]$

ii) $P(\Omega) = 1$

iii) σ-additivity

(Ω, F, P) is called a probability space.

1.3 Null sets

Let (Ω, F, μ) be a measure space.

Every $N \in F$ where $\mu(N) = 0$ is a null set

1.4 Construction of measures

Idea: Functions with properties as measures (premeasures defined on a ring) can be extended to complete measures on the σ-algebra generated by the ring.

Definition 1.18: Ring

$R \in p(\Omega)$ is a ring on Ω if:

i) $\phi \in R$

ii) $A, B \in R \implies A \backslash B \in R$

iii) $A, B \in R \implies A \cup B \in R$

A premeasure μ_0 on R is a function such that:

i) $\mu_0 : R \to [0, \infty]$

ii) $\mu(\phi) = 0$

iii) σ-additivity

Theorem 1.19: Caratheorody's exclusion theorem

Let μ_0 be a premeasure to R on Ω. There exists a complete measure μ on $F : \sigma(R)$ which coincides with μ_0 on R.

If μ_0 is σ-finite, then μ is unique.

Theorem 1.21

$F : R \to R$ right-continuous and increasing. There exists a unique Borel measure μ_F such that $\mu_F([a, b]) = F(b) - F(a)$.

Remark 1.22

1) By Theorem 1.19, μ_F is complete and is called the Lebesgue-Stielties measure associated to F. It's domain $\overline{B(R)}$ known as Lebesgue σ-algebra can be shown to strictly contain $B(R)$. Sets in $\overline{B(R)}$ are Lebesgue measurable (or Lebesgue sets)

2) If $F(x) = x$, $\Lambda := \mu_F$ is called Lebesgue measure on R and sets $N \in \overline{B(R)} : \Lambda(N) = 0$ Lebesgue null sets.

Remark 1.24

1) Theorem 1.21 extends to $F : R^d \to R$ which is:

i) Right-continuous: $F(\underline{x}) = lim_{\underline{h} \to 0} F(\underline{x} + \underline{h})$

ii) d-increasing: The F-volume $\Delta_{(\underline{a}, \underline{b}]} F$ of $(\underline{a}, \underline{b}]$ is ≥ 0 for all $\underline{a} \leq \underline{b}$, where:

$$\Delta_{[\underline{a}, \underline{b}]} F = \Pi_{i=1}^{d} (b_j - a_j)$$

$$\text{e.g. } d = 2, \ \underline{a} = (a_1, a_2), \underline{b} = (b_1, b_2)$$

$$\Delta_{[\underline{a}, \underline{b}]} F = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

2) If $lim_{x_j \to -\infty} F(\underline{x}) = 0$ for some $j \in \{1, \ldots, d\}$ and $F(\infty) = lim_{\underline{x} \to -\infty} F(\underline{x}) = 1$, then μ_F is a probability measure on $B(R^d)$.

2. Geometric and Laplace Probability

Prop 2.1

Let (Ω, F, μ) be a measure space, where $0 < \mu(\Omega) < \infty$

Then (Ω, F, P) with $P(A) = \frac{\mu(A)}{\mu(\Omega)}$ is a probability space.

F is a σ-algebra on Ω and $\Omega' \subseteq \Omega$, then one can show: The restriction $F|_{\Omega} := \{A \cup \Omega' : A \in F\}$ is a σ-algebra on Ω'.

Ref 2.2

$\Omega \subseteq R^d : 0 < \Lambda(\Omega) < \infty, F = \overline{B}(\Omega), p(A) = \frac{\mu(A)}{\mu(\Omega)}$, for all $A \in F$, then the probability space (Ω, F, P) is called geometric probability space.

Prop 2.4

$1 \leq |\Omega| < \infty,\ F = P(\Omega),\ P(A) = \frac{|A|}{|\Omega|}$

Then (Ω, F, P) is a finite probabiity space called Laplace probability space.

P is called discrete uniform distribution on Ω.
Remark 2.5

For Laplace probability space, the probability mass function on Ω is:

$$f(w) = P(\{\omega\}) = \frac{|\{\omega\}|}{|\Omega|} = \frac{1}{|\Omega|}, \forall \omega \in \Omega$$

So the discrete uniform distribution assigns equal probability $\frac{1}{|\Omega|}$ to each $\omega \in \Omega$

3. Probability counting techniques

3.1 Basic rules

Prop 3.1

1) Addition rule: If $A_1, \dots, A_n$ are pairwise disjoint finite sets, then:

$$| \cup_{i=1}^{n} A_i| = \Sigma_{i=1}^{n} |A_i|$$

2) Multiplication rule: If $A_1, \dots, A_n$ are pairwise disjoint finite sets, then:

$$|\Pi_{i=1}^{n} A_i| = |\{(\omega_1, \dots, \omega_n) : \omega_i \in A_i \forall i\} = \Pi_{i=1}^{n} |A_i|$$

3.2 Urn models

Draw k balls from n balls. Let the balls be numbered $\{1, \dots, n\}$ The 4 classical models are:

I) With order, with replacement:

$$\Omega_I = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\}\} = \{1, \dots, n\}^k$$

$$|\Omega_I| = n^k$$

II) With order, without replacement:

$$\Omega_{II} = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\}, \omega_i \neq \omega_j, \forall i \neq j\}$$

$$|\Omega_{II}| = n(n-1)\dots(n-k+1) = nPk$$

III) Without order, without replacement:

$$\Omega_{III} = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\}, \omega_1 < \dots < \omega_k\}$$

$$|\Omega_{III}| = \frac{|\Omega_{II}|}{k!} = \binom{n}{k}$$

IV) Without order, with replacement:

$$\Omega_{IV} = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\}, \omega_1 \leq \dots \leq \omega_k\}$$

$$|\Omega_{IV}| = \binom{n+k-1}{k}$$

4. Conditional Probability and Independence

Prop 4.1

Let (Ω, F, P) be a probability space. Let $B \in F$ such that $P(B) > 0$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, A \in F$$

This is a probability measure on (Ω, F), the ordinary conditional probability of A given B.

Prove: $P(A|B)$ is a probability measure:

i) $P : F \to [0, 1]$
Since $A \subseteq B$, this means $0 \leq P(A \cap B) \leq P(B)$. So $P(A|B) \in [0, 1]$

ii) $P(\Omega|B) = 1$
This is true since $\Omega \cap B = B$

iii) σ-additivity

Let $A_i \in F$ be disjoint.

$$P(\cup_{i=1}^{\infty} A_i|B) = \frac{P((\cup_{i=1}^{\infty} A_i) \cap B)}{P(B)} = \frac{P(\cup_{i=1}^{\infty}(A_i \cap B))}{P(B)}$$

Since A_i are disjoint, this means $A_i \cap B$ are disjoint.

$$\frac{P(\cup_{i=1}^{\infty}(A_i \cap B))}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)}$$

Thus, $P(A|B)$ is a probability measure.

Theorem 4.2: Law of total probability

(Ω, F, P) is a probability space. $B_1, B_2, \cdots \in F$ is a partition of Ω, that is, $\Omega = \cup_{i=1}^{\infty} B_i$. B_i are disjoint. Then:

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i)$$

Thm 4.5: Bayes' Theorem

(Ω, F, P) is a probability space. $A, B \in F$ such that $P(A) > 0, P(B) > 0$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

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

Definition 4.7

Let (Ω, F, P) be a probability space. Then:

(i) $A_1, A_2 \in F$ are independent if $P(A_1 \cap A_2) = P(A_1)P(A_2)$

(ii) $A_1, \ldots, A_n \in F$ are independent if $P(\cap_{i \in I} A_i) = \Pi_{i \in I} P(A_i)$ for $\forall I \subseteq \{1, \ldots, n\}$

(iii) $a_1, \ldots, a_n \in F$ are independent if $A_1, \ldots, A_n \in F$ are independent, $\forall A_i \in a_i$

(iv) $\{A_i\} \in F$ are independent if $a_1, \ldots, a_n \in F$ are independent

5. Random Variables and Distributions

Let (Ω, F, P) be a probability space. When predicting outcomes of experiments, it's useful to consider mappings $X : \Omega \to \Omega'$ for some measurable set (Ω', F'), often $(R, B(R))$ or $(R^d, B(R^d))$

Definition 5.1

The preimage of $X : \Omega \to \Omega'$ is defined by $X^{-1}(A') = \{w \in \Omega : X(w) \in A'\}, A' \subseteq \Omega'$

Lemma 5.2

(i) $X^{-1}(\phi) = \phi, X^{-1}(\Omega') = \Omega$

(ii) $(X^{-1}(A'))^c = X^{-1}((A')^c), \forall A' \subseteq \Omega'$

(iii) $\cup_{i\in I} X^{-1}(Ai') = X^{-1}(\cup_{i\in I} A'), \cap_{i\in I} X^{-1}(Ai') = X^{-1}(\cap_{i\in I} A')$

If F' is a σ-algebra on Ω', then $\sigma(X) = \{X^{-1}(A') : A' \in F'\}$ is a σ-algebra on Ω, the σ-algebra generated by X.

Definition 5.3

Let (Ω, F), (Ω', F') be measurable sets. $X : \Omega \to \Omega'$ is called (F, F')-measurable if $\sigma(X) \subseteq F$ i.e. $X^{-1}(A') \in F, \forall A' \in F'$.

If (Ω', F') is $(R, B(R))$ or $(R^d, B(R^d))$, then X is called a random variable or random vector.

Remark 5.5

(1) We typically write X (e.g. $X = 1_A$) instead of $X(w)$ (e.g. $X(w) = 1_A(w), w \in \Omega$)

(2) We can study sequences of random variables $X_1, X_2, \ldots$. Such sequences play a role in major limiting results

(3) One can show that the measurability is preserved by many operations e.g. compositions of measurable functions are measurable.

(4) Random vectors are vectors of random variables.

Prop 5.6

Let (Ω, F, P) be a probability space. Let (Ω', F') be a measurable set. If $X : \Omega \to \Omega'$ is measurable, then $P_X = P(X^{-1})$ is a probability measure on (Ω', F'), the distribution of X.

$$P_X(w) = P(X^{-1}(w)), w \in \Omega'$$

Remark 5.7

(1) P_x assigns probabilities to events involving measurable X since $X^{-1}(A') \in F, \forall A' \in F$ and P knows how to assign probabiliities to such events.

(2) We often write $P(x \in A) = P(\{w \in \Omega : X(w) \in A'\}) = P(X^{-1}(A')) = P_X(A')$

(3) If X is a random variable, then the distribution of X is a Borel probability measure on R.

$$\mathcal{F}(x) := P_x((-\infty, x]) = P(X \in (-\infty, x]) = P(\{w \in \Omega : X(w) \leq x\}), \forall X \in R$$

We call $\mathcal{F}(x)$ the distribution function (DF) of X and write $X \sim \mathcal{F}$

The following characterizes all DF on R.

Theorem 5.8

$\mathcal{F} : R \to [0,1]$ is the DF of a unique Borel probability measure P_F on $R \iff$:

(1) $\mathcal{F}(-\infty) = 0, \mathcal{F}(\infty) = 1$

(2) $a \geq b \implies F(a) \geq F(b)$

(3) $\mathcal{F}$ is right-continuous

Remark 5.9

" $\implies$ " implies properties of any DF $\mathcal{F}$.

"$\impliedby$" implies $\forall \mathcal{F}$:DF induces a unique distribution

e.g. $\mathcal{F}(x) = min(max(0,x),1), \forall x \in R$ is the DF of the standard uniform distribution U(0,1).

e.g. $\mathcal{F}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{z-\mu}{\sigma})^2} dz$ is the DF of the normal distribution $N(\mu, \sigma^2)$.

$$P(x \in A') = P(\{w \in \Omega : X(w) \in A'\}) = P(A \in F)$$

$$\mathcal{F}(x) = P(X \leq x) = P(X \in (-\infty, x]) = P(X \in A')$$

$$X \sim \mathcal{F} \iff \mathcal{F}(X) = P(X \leq x), \forall x \in R$$

Ref 5.10

(1) $\mathcal{F} : R \to R$ is monotone increasing. The generalized inverse $\mathcal{F}^-$ of $\mathcal{F}$ is defined by:

$$\mathcal{F}^-(y) = inf\{x \in R : F(x) \geq y\}, y \in (0,1)$$

(2) If $\mathcal{F}$ is a DF, then $\mathcal{F}^-$ is the quantile function of $\mathcal{F}$.

Remark 5.11

(2) Some facts:

- If $\mathcal{F}$ is increasing and continuous, then $\mathcal{F}^- = \mathcal{F}^{-1}$
- Graph of of $\mathcal{F}^-$ is obtained by mirroring $\mathcal{F}$ through $y = x$
- $\mathcal{F}^-$ is increasing and left-continuous
- One can work with $\mathcal{F}^-$ as with $\mathcal{F}^{-1}$ but be careful: If $\mathcal{F}$ is discontinous at $x = a$, i.e. $\lim_{x \to a-} \mathcal{F}(x) = b$ and $\lim_{x \to a+} \mathcal{F}(x) = c$, then $\mathcal{F}(F^-(x)) = c, \forall x \in [b,c]$

Prop 5.12

If $F : R \to [0,1]$: satisfies i)-iii) of Theorem 5.8, then there exists a probability space (Ω, F, P) and a random variable $X : \Omega \to R$ such that $X \sim \mathcal{F}$

Prop 5.12 can be exploited for generating (pseudo-)random numbers, that is, numbers which reasonable realizations of $X \sim \mathcal{F}$ on a computer based on the following result known as inversion method for sampling. Note that there are various ways to sample from $U(0,1)$ (with DF $\mathcal{F}_U(x) = x, \forall x \in [0,1]$) on the computer.

Prop 5.13

Let $\mathcal{F}$ be a DF. $U \sim U(0,1)$. $X := \mathcal{F}^{-}(U) \sim \mathcal{F}$.

Remark 5.14

(1) $X \sim \mathcal{F}$. $P_x((a,b]) = P_{\mathcal{F}}((a,b]) = \mathcal{F}(b) - \mathcal{F}(a)$

(2) $X \sim \mathcal{F}$. Then $P(X = x) = P(X \in \cap_{n=1}^{\infty}(x - \frac{1}{n}, x])$
$= \lim_{n\to\infty} P(X \in (x - \frac{1}{n}, x])$
$= \mathcal{F}(x) - \lim_{n\to\infty} \mathcal{F}(x - \frac{1}{n})$
$= \mathcal{F}(x) - \mathcal{F}(x^{-}), \forall x \in R$

- If $\mathcal{F}$ is continuous in x, then $\mathcal{F}(x-) = \mathcal{F}(x)$ i.e. $P(X = x) = 0$
- If $\mathcal{F}$ is constant in $(a,b]$, then $P(x \in (a,b]) = \mathcal{F}(b) - \mathcal{F}(a) = 0$ so $X \sim \mathcal{F}$ does not take values in $(a,b]$
- If $\mathcal{F}$ jumps in x, then $P(X = x) = \mathcal{F}(x+) - \mathcal{F}(x-) =$ Jump height of $\mathcal{F}$ in x
- Note: In each jump "gap" $(\mathcal{F}(x+), \mathcal{F}(x-)]$, there exists a rational number. And since $\mathcal{F}$ is increasing, they are all distinct. So: Number of jumps $\leq |Q|$ so DFs can at most countable many jumps.

Definition 5.15

Let X be a random variable with distribution Px and DF $\mathcal{F}$.

(1) X is a discrete random variable if $\exists S = \{x_1, \dots\} \subseteq R : P(X \in S) = 1$. $\mathcal{F}$ is a step function. $ran(\mathcal{F}) = \{\mathcal{F}(x) : x \in S\}$ is at most countable. $f(x) = P(X = x)$ is called probability mass function (PMF) of $X(P_x, \mathcal{F})$.

(2) $X(P_x, \mathcal{F})$ is a continuous random variable if $\mathcal{F}$ is continuous.

(3) $X(P_x, \mathcal{F})$ is an absolutely continuous random variable if $\mathcal{F}(x) = \int_{-\infty}^{x} f(z)dz, \forall x \in R$ for some integrable $f : R \to [0, \infty)$ such that $\int_{-\infty}^{\infty} f(z)dz = 1$. f is called density of $X(P_x, \mathcal{F})$.

Remark 5.16

(1) DFs $\mathcal{F}$ could be discrete and continuous/absolutely continuous. This is called mixed-type DF.

(2) $\mathcal{F}$ is differentiable on R and $\mathcal{F}'$ is integrable $\implies$ $\mathcal{F}$ is absolutely continuous with density $f = \mathcal{F}'$. Absolutely continuous $\implies$ continuous, since:

$$|\mathcal{F}(x+h) - \mathcal{F}(x)| = |\int_{x}^{x+h} f(z)dz| \leq \int_{x}^{x+h} |f(z)|dz \leq sup\{|f(z)|\} \int_{x}^{x+h} dx = sup\{|f(z)|\}h$$

$$\lim_{h\to 0} sup\{|f(z)|\}h = 0$$

Thus, $X \sim \mathcal{F} \implies P(X = x) = 0, \forall x \in R$. So $P(X = x) \neq f(x)$ in general. However:

$$P(X \in (x - \epsilon]) = \int_{x-\epsilon}^{x} f(z)dz \approx \epsilon f(x)$$

for small $\epsilon > 0$.

(3) Not all continuous $\mathcal{F}$ are absolutely continuous. Example: Cantor set.

(4) For PMFs on densities, always provide a domain. e.g. $f(x) = \frac{1}{c}$ is a density on $[0, c]$ but not $[0, \infty)$.

Definition 5.18

Let (Ω, F, P) be a probability space. $\underline{X} = (X_1, \dots, X_j)$ is a random vector. Then $P_{\underline{X}} = P \circ \underline{X}^{-1}$. $P_{\underline{X}}(B) = P(\underline{X}^{-1}(B)) = P(\{w \in \Omega : \underline{X}(w) \in B\}), \forall B \in B(R^d)$ is the distribution of $\underline{X}$ and:

$$\mathcal{F}(\underline{X}) := P_{\underline{X}}((\underline{-\infty}, \underline{x}]) = P(\{w \in \Omega : X_j(w) \leq x_j, \forall j = 1, \dots, d\})$$

$\underline{X} \sim \mathcal{F}$: DF of $\underline{X}$

$$\underline{X} \sim \mathcal{F} \iff \mathcal{F}(\underline{x}) = P(\underline{X} \leq \underline{x}), \forall \underline{x} \in R^d$$

We call $F_j(x_j) = \mathcal{F}(\infty, \dots, \infty, x_j, \infty, \dots, \infty)$ the jth margin of $\mathcal{F}$ or the jth marginal DF of $\underline{X}$.

Theorem 5.19

$\mathcal{F} : R^d \to [0,1]$ is the DF of a unique Borel probability measure $P_{\mathcal{F}}(= P_x)$ on R^d iff:

(1) $\lim_{x_j \to -\infty} F(\underline{X}) = 0$ for any $j \in 1, \dots, d$, and $\lim_{x \to \infty} F(\underline{X}) = 1$

(2) $\mathcal{F}$ is d-increasing: $\mathcal{F}(\underline{b}) \geq \mathcal{F}(\underline{a}), \forall \underline{b} \geq \underline{a}$

(3) $\mathcal{F}$ is right-continuous: $\lim_{\underline{h} \to \underline{0}} \mathcal{F}(\underline{x} + \underline{h}) = \mathcal{F}(\underline{x})$
    If $\mathcal{F}$ is a DF and $X \sim \mathcal{F}$, then:

$$\Delta_{(\underline{a}, \underline{b})} \mathcal{F} = \mu_{\mathcal{F}}(\underline{a}, \underline{b}] = P_{\underline{X}}((\underline{a}, \underline{b}]) = P(X \in (\underline{a}, \underline{b}])$$

Definition 5.20

$\underline{X} \sim \mathcal{F}$ with distribution $P_{\underline{X}}$, then:

(1) $\underline{X}(P_{\underline{X}}, \mathcal{F})$ is discrete if $\exists \lambda = \{\underline{X_1}, \dots\} : P(\underline{X} \in \lambda) = 1$.

    i.e. $\mathcal{F}$ is a step function: $Ran(\mathcal{F})$ is countable.

    Then $f(\underline{x_i}) = P(\underline{X} = \underline{x_i})$ is PMF of $\underline{X}(P_{\underline{X}}, \mathcal{F})$

(2) $\underline{X}(P_{\underline{X}}, \mathcal{F})$ is continuous if $\mathcal{F}$ is continuous

(3) $\underline{X}(P_{\underline{X}}, \mathcal{F})$ is absolutely continuous if $\mathcal{F}(\underline{x}) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f dx_1 \dots dx_n$ for some integrable f.

Remark 5.21

(1) If f is absolutely continuous, then $f_j(x_j) = \frac{d}{dx_j} \mathcal{F}_j(x_j)$ is the jth marginal density of $\mathcal{F}$.

$$f_j(x_j) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_{j+1}} \int_{-\infty}^{x_{j-1}} \cdots \int_{-\infty}^{x_1} f dx_d \dots x_{j+1} x_{j-1} \dots x_1$$

(2) f is absolutely continuous $\implies$ margins of f are absolutely continuous.

6. Independence and Dependence

Definition 6.1

Let (Ω, F, P) be a probability space. Let (Ω', F') be a measurable space. Let $X_i : \Omega \to \Omega'$ be (F, F')-measurable, $\forall i \in I \subseteq R$. Then $X_{i, i\in I}$ are independent if $\sigma(x_i), i \in I$ are independent, that is, $X_{i_1}^{-1}(A'_{i_1}), \dots, X_{i_n}^{-1}(A'_{i_n})$ are independent for $A_{i_1}, \dots, A_{i_n} \in F, i_1, \dots, i_n \subseteq I, n \in N$.

Remark 6.2

(1) $X_1, \dots, X_d \sim \mathcal{F}$ are independent

$$\iff X_1^{-1}(B_1), \dots, X_d^{-1}(B_d) \text{ are independent, } \forall B_1, \dots, B_d \in B(R).$$

$$\iff P(\cap_{j\in\delta} X_j^{-1}(B_j)) = \prod_{j\in\delta} P(X_j^{-1}(B_j)), \forall \delta \in \{1, \dots, d\}$$

It suffices to consider: $B_j = \{(-\infty, x] : x \in R\}, j = \{1, \dots, d\}$

(2) If $\mathcal{F}$ is absolutely continuous, then $X_1, \dots, X_d$ are independent

$$\iff f(\underline{x}) = \frac{\partial^d}{\partial x_d \dots \partial x_1} \mathcal{F}(\underline{x}) = \prod_{j=1}^{d} \frac{\partial}{\partial x_i} \mathcal{F}_j(x_j) = \prod_{j=1}^{d} f_j(x_j), \forall \underline{x} \in R^d$$

If $\mathcal{F}$ is discrete, then $X_1, \dots, X_d$ are independent with support $S = \{\underline{X}_1, \dots\}$

$$\iff f(\underline{x}) = \frac{\partial^d}{\partial x_d \dots \partial x_1} \mathcal{F}(\underline{x}) = \prod_{j=1}^{d} \frac{\partial}{\partial x_i} \mathcal{F}_j(x_j) = \prod_{j=1}^{d} f_j(x_j), \forall \underline{x} \in S$$

(3) If $X_{j_1}, \dots, X - j_d$ are independent random variables and $h_j : R^{d_j} \to R$ are $(B(R^{d_j}), B(R))$-measurable, then $Y_j = h_j(X_{j_{d_0}}, \dots, X_{j_{d_j}}), j \in N$ are also independent random variables.

i.e. measurable functions of independent random variables are independent random variables

Prop 6.4

For DFs $\mathcal{F}_1, \dots, \mathcal{F}_d$, there exists a probability space (Ω, F, P) and independent random variables $X_1, \dots, X_d$ such that $X_j \sim F_j, \forall j \in 1, \dots, d$

Definition 6.5

A copula C is a DF with $U(0,1)$ margins

Prop 6.6

$C : [0,1]^d \to [0,1]$ is a d-dimensional copula iff:

(1) $C((\underline{u})) = 0 \impliedby u_j = 0$ for at least one $j \in \{1, \dots, d\}$ (Groundedness)

(2) $C(1, \dots, 1, u_j, 1 \dots, 1) = u_j$

(3) $\Delta_{(\underline{a}, \underline{b}]} C \geq 0, \forall \underline{a}, \underline{b} \in [0,1]^d : \underline{a} \geq \underline{b}$ (d-increasing)

Theorem 6.8

(1) For all DFs $\mathcal{F}$ with continuous margins $\mathcal{F}_1, \ldots, \mathcal{F}_d$, there exists a unique copula C:

$$\mathcal{F}(\underline{x}) = C(\mathcal{F}_1(x_1), \ldots, \mathcal{F}_d(x_d)), \forall \underline{x} \in R^d$$

Given by:

$$C(\underline{u}) = \mathcal{F}(\mathcal{F}_1^{-}(x_1), \ldots, \mathcal{F}_d^{-}(x_d)), \forall \underline{u} \in [0,1]^d$$

(2) Given a copula C and univariate DFs $\mathcal{F}_1, \ldots, \mathcal{F}_d$, $\mathcal{F}$ defined by $\mathcal{F}(\underline{x}) = C(\mathcal{F}_1(x_1), \ldots, \mathcal{F}_d(x_d)), \forall \underline{x} \in R^d$ is a DF with margins $\mathcal{F}_1, \ldots, \mathcal{F}_d$.

Remark 6.9

(1) Any continuous function $\mathcal{F}$ can be decomposed into its copula. The copula is thus, the function that contains information about the dependence between its random variables.

(2) 2 important results about copulas:

- Frechet-Hoeffding bounds: For copula C:

$$W(\underline{u}) \leq C(\underline{u}) \leq M(\underline{u}), \forall \underline{u} \in [0,1]^d$$

- Invariance principle: If $T_1, \ldots, T_d$ are strictly increasing, then $(T_1(X_1), \ldots, T_d(X_d))$ also has copula C.

7. Summary Statistics

7.1 Expectation and Moments

Definition 7.1

Let (Ω, F, P) be a probability space with random variable X.

If $E(X^+) < \infty$ or $E(X^-) < \infty$, then X is called quasi-integrable and $E(X) = \int_\Omega X dP$.

If $E(|X|) < \infty$ (notation, $X \in L^1(\Omega, F, P)$), X is integrable and $E(X)$: Expectation of X.

Remark 7.2 (Change of variables)

Let (Ω, F, P) be a probability space with random variable $X \sim \mathcal{F}$. $h : R \to R$ is measurable, and $E(|h(X)|) < \infty$, then:

$$E(h(X)) = \int_\Omega h(X)dP = \int_R h(x)d\mathcal{F}(x)$$

(If discrete) $E(h(X)) = \sum_{x \in S} h(x)f(x)$

(If absolutely continuous) $E(h(X)) = \int_R h(x)f(x)dx$

Remark 7.3

(1) $E(|h(X)|) < \infty \iff \int_R |h(x)|d\mathcal{F}(x) < \infty$

(2) $E(h(X))$ can be computed by:

- $\int_R h(X)d\mathcal{F}_X(x) < \infty$. $\mathcal{F}_X$: DF of X
- $Y = h(X) \;\to\; \int_R Y d\mathcal{F}_Y(y)$. F_Y: DF of Y

(3) $h(X) = X^p, p > 0$, the pth amount of $X \sim \mathcal{F}$. $E(X^p) = \int_R X^p d\mathcal{F}_X(x) < \infty$.

With Koldon's inequality, $q < p$ and $E(|X|^p) < \infty \implies E(|X|^q) < \infty$.

i.e. Existing higher moments $\implies$ Existing lower moments

(4) Theorem 7.2 extends to random vectors $\underline{X} \sim \mathcal{F}$ and measurable $h : R^d \to R$ such that $E(|h(\underline{X})|) < \infty$.

Then $E(h(\underline{X})) = \int_{R^d} h(\underline{X})d\mathcal{F}(\underline{X})$

Prop 7.4

$X, Y \in L'(\Omega, F, P)$. Then:

(1) $aX + bY \in L^1(\Omega, F, P)$ with $E(aX + bY) = aE(X) + bE(Y)$ (Linearity)

(2) $A \in F \implies E(1_A) = P(A)$

(3) $X \geq 0$ a.s. $\implies E(X) \geq 0$

(4) $X \leq Y$ a.s. $\implies E(X) \leq E(Y)$ (monotinicity)

(5) $X \geq 0$ a.s. $E(X) = 0 \implies X = 0$ a.s.

(6) $|E(X)| \leq E(|X|)$

(7) $X \in \mathcal{F}$, then:

- $E(X) = \int_0^\infty 1 - \mathcal{F}(X)dx - \int_{-\infty}^0 \mathcal{F}(X)dx$

- $E(X) = \int_0^1 \mathcal{F}^-(u)du$

Prop 7.5

If X is quasi-integrable, with $P(A) = 0$, then

$$\int_A XdP := \int_\Omega X1_A dP = E(X1_A) = 0$$

7.2 Variance and Covariance

Definition 7.6

Let $(\Omega, \mathcal{F}, P)$ be a probability space with random variables X, Y such that $E(X^2) < \infty, E(Y^2) < \infty$ (notation: $X, Y \in L^2(\Omega, \mathcal{F}, P)$). Then,

$$Var(X) = E((X - EX)^2)$$

is called the variance of X (or its distribution or df), and $\sqrt{Var(X)}$ is the standard deviation.

The covariance of X, Y is defined by

$$Cov(X, Y) := E((X - EX)(Y - EY))$$

and the correlation of X, Y by

$$Cor(X, Y) := \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Remark 7.7

Here are some properties of variance and covariance.

1.

$$\begin{aligned} Var(X) &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Similarly,

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Also note that

$$\begin{gathered} Cov(X,X) = Var(X) \\ Cov(X,c) = 0 \ \forall c \in R \\ Cov(X,Y) = Cov(Y,X) \end{gathered}$$

2.

$$\begin{aligned} Var(X) = 0 &\iff E((X - EX)^2) = 0 \\ &\iff (X - EX)^2 = 0 \text{ almost surely} \\ &\iff X - EX = 0 \text{ almost surely} \\ &\iff X = EX \text{ almost surely} \end{aligned}$$

3.

$$\begin{aligned} Var(aX + bY) &= E(((aX + bY) - E(aX + bY))^2) \\ &= E((a(X - EX))^2) + 2E(a(X - EX)b(Y - EY)) + E((b(Y - EY))^2) \\ &= a^2 Var(X) + 2abCov(X,Y) + B^2 Var(Y) \end{aligned}$$

for $a, b \in R$. In particular if $Y = 1$ almost surely, $Var(aX + b) = a^2 Var(X)$.

4. If X, Y are independent,

$$E(XY) = E(X)E(Y)$$

which implies

$$Cov(X,Y) = 0 = Cor(X,Y)$$

So, independence implies uncorrelatedness. The converse is not true in general:

Let $X \sim U(-1,1), Y = X^2$. Then,

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= E(X^3) - 0E(X^2) \\ &= 0 \end{aligned}$$

by symmetry, since X^3 is an odd function. So X, Y are uncorrelated but dependent.

Prop 7.8

Let $X, Y \in L^2(\Omega, \mathcal{F}, P)$. Then,

$$\begin{aligned} &\rho := Cor(X,Y) \in [-1,1] \text{ and } |\rho| = 1 \\ \iff &Y \text{ is a linear function of } X \text{ with slope } \lesseqgtr 0 \\ \iff &\rho = \pm 1 \end{aligned}$$

8. Examples of Distributions

8.1 Discrete Distributions

8.1.1 Discrete Uniform Distribution

Notation: $U(\{x_1, \dots, x_n\})$

$X \sim U(\{x_1, \dots, x_n\})$ models n distinct outcomes, each with equal probability

PMF:

$$f(x) = \begin{cases} \frac{1}{n} & x \in \{x_1, \dots, x_n\} \\ 0 & \text{otherwise} \end{cases}$$

Check: $f(x) \geq 0$ for $x \in R$, and $\sum_{i=1}^{n} f(x_n) = \frac{1}{n}n = 1$

Distribution function:

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \sum_{k \in \{1, \dots, n\}: x_k \leq x} P(X = x_k) \\ &= \frac{1}{n}|\{x_k : x_k \leq x\}| \end{aligned}$$

If $x_k = k$, we have

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{\lfloor x \rfloor}{n} & x \in [1, n) \\ 1 & x \geq n \end{cases}$$

Mean:

$$E(X) = \sum_{k=1}^{n} x_k P(X = x_k) = \frac{1}{n} \sum_{k=1}^{n} x_k$$

.
If $x_k = k$, then

$$E(X) = \frac{n+1}{2}$$

.

Variance:

$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{n} \sum_{k=1}^{n} x_k^2 - (\frac{1}{n} \sum_{k=1}^{n} x_k)^2$$

.
If $x_k = k$, then

$$Var(X) = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} - (\frac{n+1}{2})^2 = \frac{n^2 - 1}{12}$$

If $U \sim U(0,1)$, then $\lceil nU \rceil \sim U(\{1, \dots, n\})$ since

$$P(nU = k) = P(k - 1 < nU \leq k) = P(U \in (\frac{k-1}{n}, \frac{k}{n}]) = \frac{k}{n} - \frac{k-1}{n} = \frac{1}{n}$$

for $k = 1, \dots, n$. In particular, $X = x_{\lceil nU \rceil} \sim U(\{1, \dots, n\})$

8.1.2 Binomial Distribution

Notation: $B(n,p)$ where $n \geq 1, p \in (0,1)$

$X \sim B(n,p)$ models the number of successes when independently repeating same experiment with outcomes success or failure n times, where $P(\text{success}) = p$. These experiments are Bernoulli trials. Then, $X = \sum_{k=1}^{n} X_k$ for $X_1, \ldots, X_n \sim B(1,p)$.

PMF:

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \{0, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$$

Check: $f(x) \geq 0$ for $x \in R$, and

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

Distribution function:

$$F(X) = P(X \leq x) = P(X \leq \lfloor x \rfloor) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{x} p^k (1-p)^{n-k} = \int_p^1 f_{x,n}(z) dz$$

where $f_{x,n}$ is the density of the $\text{Beta}(x+1, n-x)$ distribution. By letting $p=0$, this gives 1 for $x \in [0,n]$.

Mean:

$$E(X) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-k} = np$$

Variance:

$$\begin{aligned} E(X^2) - E(X) &= E(X(X-1)) \\ &= \sum_{k=2}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= p^2 n(n-1) \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{n-k} \\ &= p^2 n(n-1) \\ \Rightarrow E(X^2) &= p^2 n(n-1) + np \\ Var(X) &= p^2 n(n-1) + np - (np)^2 \\ &= np(1-p) \end{aligned}$$

8.1.3 Geometric Distribution

Notation: $Geo(p), p \in (0,1)$

$X \sim Geo(p)$ models the number of independent Bernoulli trials with success probability p until first success.

PMF:

$$f(x) = \begin{cases} p(1-p)^{x-1} & x \in N \\ 0 & \text{otherwise} \end{cases}$$

Check: $f(x) \geq 0$ for $x \in R$. $\sum_{k=1}^{\infty} p(1-p)^{k-1} = p\sum_{k=0}^{\infty}(1-p)^{k-1} = p\frac{1}{1-(1-p)} = 1$

Distribution function:

$$\begin{aligned} F(x) &= P(X \leq \lfloor x \rfloor) \\ &= \sum_{k=1}^{\lfloor x \rfloor} p(1-p)^{k-1} \\ &= p\sum_{k=0}^{\lfloor x-1 \rfloor} p(1-p)^{k-1+1} \\ &= p\frac{1-(1-p)^{\lfloor x \rfloor}}{1-(1-p)} \\ &= 1-(1-p)^{\lfloor x \rfloor} \end{aligned}$$

for $x \in [1, \infty)$.

Mean: Note that

$$\sum_{k=1}^{\infty} kq^{k-1} = \frac{d}{dq}\sum_{k=1}^{\infty} q^k = \frac{d}{dq}(\frac{1}{1-q} - 1) = \frac{1}{(1-q)^2}$$

for $|q| < 1$. Therefore,

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p\frac{1}{1-(1-p)^2} = \frac{1}{p}$$

Variance: Note that

$$\sum_{k=2}^{\infty} k(k-1)q^{k-2} = \sum_{k=0}^{\infty} \frac{d^2}{dq^2}q^k = \frac{d^2}{dq^2}(\frac{1}{1-q} - q - 1) = \frac{2}{(1-q)^3}$$

for $|q| < 1$. Therefore,

$$E(X(X-1)) = \sum_{k=2}^{\infty} k(k-1)p(1-p)^{k-1} = p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = p(1-p)(\frac{2}{(1-(1-p))^3}) = \frac{2(1-p)}{p^2}$$

$$\therefore E(X^2) = \frac{2(1-p)}{p^2} + E(X) = \frac{2-p}{p^2}$$

$$\therefore Var(X) = \frac{1-p}{p^2}$$

8.1.4 Poisson Distribution

Notation: $Poi(\lambda), \lambda > 0$

$X \sim Poi(\lambda)$ models the number of events occurring in fixed time interval, if these events occur at fixed rate λ and independently of time of last event.

PMF:

$$f(x) = \begin{cases} \frac{\lambda^x}{x!}e^{-\lambda} & x \in N_0 \\ 0 & \text{otherwise} \end{cases}$$

Check: $f(x) \geq 0$ for $x \in R$. $\sum_{x=0}^{\infty} f(x) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} = e^{\lambda}e^{-\lambda} = 1$

Distribution function:

$$F(x) = \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k}{k!}e^{-\lambda}$$

for $x \in R$. R uses $F(x) = \frac{\Gamma(\lfloor x \rfloor + 1, \lambda)}{\lfloor x \rfloor !}$, where $\Gamma(s, z) = \int_z^{\infty} t^{s-1}e^{-t}dt$ is the upper incomplete gamma function.

Mean:

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

Variance:

$$\begin{aligned} E(X(X-1)) &= \sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} = \lambda^2 \\ &\Rightarrow E(X^2) = \lambda^2 + \lambda \\ &\Rightarrow Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

8.2 Absolutely Continuous Distributions

8.2.1 Continuous Uniform Distribution

Notation: $U(a,b)$ for $a,b \in R, a < b$.

$X \sim U(a,b)$ models outcomes uniformly distributed over (a,b), that is, $P(X \in (x, x+h])$ is constant and equals $\frac{h}{b-a}$ for all $x \in [a, b-h]$.

Density:

$$f(x) = \frac{1}{b-a} 1_{(a,b]}(x)$$

for $x \in R$.

Check: $f(x) \geq 0$ for $x \in R$. Also, $\int_a^b \frac{1}{b-a} dz = \frac{1}{b-a} \int_a^b (1) dz = \frac{b-a}{b-a} = 1$.

Distribution function:

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f(z) dz \\ &= \int_a^x \frac{1}{b-a} dz \\ &= \frac{x-a}{b-a} \end{aligned}$$

for $x \in [a,b]$.

Moments:

$$\begin{aligned} E(X^k) &= \int_a^b x^k f(x) dx = \frac{1}{b-a} \int_a^b x^k dx = \frac{b^{k+1} - a^{k+1}}{(b-a)(k+1)} = \frac{\sum_{l=0}^{k} a^l b^{k-l}}{k+1} \\ E(X) &= \frac{b-a}{2} \\ Var(X) &= \frac{(b-a)^2}{12} \end{aligned}$$

8.2.2 Gamma Distribution

Notation: $\Gamma(\alpha, \beta)$ where $\alpha > 0$ is the shape, $\beta > 0$ is the rate.

Special cases:

- Exponential distribution
- Erlang distribution
- Chi-squared distribution

Density:

$$f(x) = \frac{\beta^x}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

for $x > 0$ where

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

is the gamma function.

Check: $f(x) \geq 0$ for $x \in R$. Also,

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z} dz = \frac{1}{\Gamma(\alpha)} \int_0^\infty \beta^\alpha (\frac{t}{\beta})^{\alpha-1} e^{-t} \frac{1}{\beta} dt = \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1$$

(Let $t = \beta z$)

Distribution function:

$$\begin{aligned} F(x) &= \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta z} z^{\alpha-1} dz \\ &= \frac{1}{\Gamma(\alpha)} \int_0^{\beta x} t^{\alpha-1} e^{-t} dt \text{ Let } t = \beta z \\ &= \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)} \end{aligned}$$

where γ is the lower incomplete gamma function (available numerically).

Moments:

$$\begin{aligned} E(X^k) &= \frac{\beta^\alpha}{\Gamma(x)} \int_0^\infty x^{k-\alpha-1} e^{-\beta x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{k(k+\alpha)}{\beta^{k+\alpha}} \int_0^\infty \frac{\beta^{k+\alpha}}{\Gamma(k+\alpha} x^{k+\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^{-k}\Gamma(k+\alpha)}{\Gamma(\alpha)} = \beta^{-k} \frac{(k_\alpha - 1) \cdot \dots \cdot (\alpha\Gamma(\alpha))}{\Gamma(\alpha)} = \beta^{-k} \prod_{i=0}^{k-1} (i+\alpha) \\ E(X) &= \frac{\alpha}{\beta} \\ Var(X) &= \frac{\alpha(1+\alpha)}{\beta^2} - (\frac{\alpha}{\beta})^2 = \frac{\alpha}{\beta^2} \end{aligned}$$

8.2.3 Exponential Distribution

Notation: $Exp(\lambda)$, $\lambda > 0$ (rate)

$X \sim Exp(\lambda)$ describes interarrival times between events in a (homogeneous) Poisson (point) process with intensity $\lambda > 0$, that is, a sequence of random variables $(N_t)_{t \geq 0}$ such that:

- $N_0 = 0$
- $\forall n \in N$ and $0 \leq t_0 < \cdots < t_n < \infty$, the increments $N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent
- $N_t - N_s \sim Poi(\lambda(t-s))$ for $0 \leq s < t$ for some $\lambda > 0$.

Such continuous-time stochastic processes model the numebr of events in a process in which events occur continuously, independently at a constant rate $\lambda > 0$ per unit (here, time) interval. Note that $N_t - N_s = N_{ts} - N_0 = N_{ts}$ for $0 \leq s < t$.

Density:

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$.

Check: $f(x) \geq 0$ for $x \in R$. Also, $\int_0^\infty \lambda e^{-\lambda x} dx = 1$. Note that $f(x)$ is $\Gamma(1, x)$ density, so $Exp(\lambda) = \Gamma(1, \lambda)$.

Distribution function:

$$F(x) = \int_0^x \lambda e^{-\lambda z} dz = 1 - e^{\lambda x}$$

for $x \geq 0$.

Moments:

$$\begin{aligned} E(X^k) &= \lambda^{-k} \prod_{i=0}^{k-1} (i+1) \\ &= \frac{k!}{\lambda^k} \\ E(X) &= \frac{1}{\lambda} \\ Var(X) &= \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2} \end{aligned}$$

8.2.4 Normal Distribution

Notation: $N(\mu, \sigma^2)$ where $\mu \in R$ is the mean/location, and $\sigma > 0$ is the standard deviation/scale.

$X \sim N(\mu, \sigma^2)$ models outcomes which fluctuate symmetrically around μ with variance σ^2.

Density:

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

for $x \in R$.

Mean:

$$E(X) = \mu$$

Variance:

$$Var(X) = \sigma^2$$

8.3 Multivariate Distributions

8.3.1 Mean vector, Covariance and Correlation Matrices

Definition 8.1

Let $\underline{X} = (X_1, \ldots, X_d)$. If $E(|X_j|) \leq \infty$ for all j, the mean vector or expectation of $\underline{X}$(or its distribution function or distribution) is defined

$$\underline{\mu} = E(\underline{X}) = (E(X_1), \ldots, E(X_d))$$

If $E(X_j) < \infty$ for all j, the covariance and correlation matrices is defined by

$$\Sigma = Cov(\underline{X}) = (Cov(X_i, X_j))_{i,j=1,\ldots,d}$$
$$P = Cor(\underline{X}) = (Cor(X_i, X_j))_{i,j=1,\ldots,d}$$

Lemma 8.2

1. $E(A\underline{X} + \underline{b}) = AE(\underline{X}) + \underline{b}$
   $E(\underline{a^T}\,\underline{X}) = \underline{a^T}E(\underline{X})$.
2. $Cov(A\underline{X} + \underline{b}) = ACov(\underline{X})A^T$
   $Var(\underline{a^T}\,\underline{X}) = Cor(\underline{a^T}\,\underline{X}) = \underline{a^T}Cov(X)\underline{a}$.

Prop 8.3

A real, symmetric matrix Σ is a covariance matrix iff Σ is positive semidefinite.

8.3.2 Normal distribution

Notation: $N(\underline{\mu}, \Sigma)$ for $\underline{\mu} \in R^d, \Sigma \in R^{d\times d}$, a covariance matrix.

$X \sim N(\underline{\mu}, \Sigma) \Leftrightarrow X = \underline{\mu} + A\underline{Z}$ where A is the Cholesky factor of Σ and $\underline{Z} = (Z_1, \ldots, Z_d)$ for $Z_j \overset{ind.}{\sim}$ for $j = 1, \ldots, d$. In other words, $\underline{X}$ is a linear transform of independent standard normal random variables. $\underline{X}$ models outcomes which fluctuate around $\underline{\mu}$ with covariance matrix Σ.

Density:

$$f_{\underline{Z}}(\underline{z}) = \prod_{j=1}^{d} f_{Z_j}(z_j) = \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}z_j^2} = \frac{e^{\frac{-1}{2}z^T z}}{(2\pi)^{\frac{d}{2}}}$$

The density $f_{\underline{X}}(\underline{x}$ of $\underline{X} = T(\underline{Z})$ for $T(\underline{Z}) = Az + \mu$ can be determined by the density transformation theorem: If T is injective and differentiable (and therefore continuous), $|\det T'(z)| > 0$ for all z, then $\underline{X} = T(z)$. Thus,

$$f_{\underline{x}}(\underline{X}) f_{\underline{Z}}(T^{-1}(\underline{X})) \frac{1}{|\det T'(T^{-1}(\underline{x}))|}$$

for all $\underline{x} \in R^d$. With $T^{-1}(\underline{X}) = A^{-1}(\underline{x} - \underline{\mu}), T'(\underline{z}) = A$ and

$$|\det T'(T^{-1}(\underline{x}))| = |\det A| = \sqrt{(\det A)^2} = \sqrt{(\det A)(\det A^T)} = \sqrt{\det AA^T} = \sqrt{\det \Sigma}$$

So we obtain

$$\begin{aligned} f_{\underline{X}}(\underline{x}) &= \frac{1}{(2\pi)^{\frac{d}{2}}} e^{\frac{-1}{2}(A^{-1}(\underline{x}-\underline{\mu}))^T (A^T(\underline{x}-\underline{\mu}))} \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} e^{\frac{-1}{2}(\underline{x}-\underline{\mu})^T (A^{-1})^T (A^{-1})(\underline{x}-\underline{\mu})} \end{aligned}$$

for all $x \in R^d$

Distribution function: only available numerically for $d \geq 3$ with so-called randomized quasi-Monte Carlo estimation via

$$F(\underline{x}) = P(\underline{X} \leq \underline{x}) = E(1_{\{\underline{X} \leq \underline{x}\}})$$

Mean vector:

$$E(\underline{X}) = \mu + AE(\underline{Z}) = \mu$$

Covariance matrix:

$$Cov(\underline{X}) = ACov(\underline{Z})A^T = AIA^T = AA^T = \Sigma$$

9. Limit Theorems

9.1 Modes of convergence

Definition 9.1

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X, X_1, \ldots, X_n : \Omega \to R$ be random variables. Then $\{X_n\}_{n \in N}$ converges to X almost surely (notation: $X_n \underset{(n\to\infty)}{\overset{a.s.}{\to}} X$) if

$$P(\lim_{n\to\infty} X_n = X) = 1$$

.

$\{X_n\}_{n\in N}$ converges to X in probability ($X_n \underset{(n\to\infty)}{\overset{p}{\to}} X$) if

$$\forall \varepsilon > 0, \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

.

If $F, F_1, F_2, \ldots$ are distribution functions with $X_n \sim F_n$ for all $n \in N$, then X_n converges in distribution ($X_n \underset{(n\to\infty)}{\overset{d}{\to}} X$) to $X \sim F$ if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x such that F is continuous at x.

Remark 9.2

1. One can show $X_n \underset{(n\to\infty)}{\overset{a.s.}{\to}} X \Rightarrow X_n \underset{(n\to\infty)}{\overset{p}{\to}} X \Rightarrow X_n \underset{(n\to\infty)}{\overset{d}{\to}} X$. Converses do not hold in general without further conditions.
2. To each of these modes of convergence is associated a limit theorem.

9.2 Weak and Strong Laws of Large Numbers

Lemma 9.3

Let $h : [0, \infty) \to [0, \infty)$ be strictly increasing and X be a random variable such that $E(H(|X|)) < \infty$. Then

$$P(|X| \geq x) \leq \frac{E(h(|x|))}{h(x)}$$

for all $x > 0$.

For $h(x) = x$, $P(|X| \geq x) \leq \frac{E(X)}{x}$ for all $x > 0$ is called Markov's inequality. For $h(x) = x^2$, $P(|X| \geq x) \leq \frac{E(X^2)}{x^2}$ for all $x > 0$ is called Chebyshev's inequality.

Prop 9.4: Weak Law of Large Numbers

If $\{X_n\}_{n \in N}$ is a sequence of iid random variables with $\mu = EX$, and $\sigma^2 = Var(X) < \infty$, then

$$\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i \underset{(n \to \infty)}{\overset{p}{\to}} \mu$$

Theorem 9.5: Strong Law of Large Numbers

If $\{X_n\}_{n \in N}$ is a sequence of iid random variables with $\mu = E(X)$, then

$$\overline{X_n} \underset{(n \to \infty)}{\overset{a.s.}{\to}} \mu$$

9.3 Central Limit Theorem

9.3.1 Characteristic Functions

Definition 9.6

The characteristic function (cf) $\phi_{\underline{X}} : R^d \to C$ of $\underline{X} \sim F$ is defined by

$$\phi_{\underline{X}}(\underline{t}) = E(e^{i\underline{t}^T\underline{X}}), t \in R^d$$
$$\text{For } d = 1, \phi_X(t) = E(e^{itx}), t \in R$$

Remark 9.7

1. By Euler's formula $e^{ix} = \cos(x) + i\sin(x)$,

$$\phi_{\underline{X}}(\underline{t}) = E(\cos(\underline{t}^T\underline{X})) + iE(\sin(\underline{t}^T\underline{X}))$$

   Therefore, $E(|e^{i\underline{t}^T\underline{X}}|) = E(\sqrt{\cos^2(\underline{t}^T\underline{X}) + \sin^2(\underline{t}^T\underline{X})}) = 1$. In particular $\phi_{\underline{X}}$ always exists, $|\phi_{\underline{X}}| \leq 1$, $\phi_{\underline{X}}(0) = 1$. Furthermore $\phi_{\underline{X}}$ is real iff

$$\begin{aligned}\phi_{\underline{X}}(\underline{t}) &= \overline{\phi_{\underline{X}}(\underline{t})} \\ &= E(\cos(\underline{t}^T\underline{X})) - iE(\sin(\underline{t}^T\underline{X})) \\ &= \phi_{\underline{X}}(-\underline{t}) \\ &= \phi_{\underline{-X}}(\underline{t})\end{aligned}$$

   for all $\underline{t} \in R^d$. That is, if $\phi_{\underline{X}}$ is point-symmetric about $\underline{0}$, or by uniqueness, if $\underline{X} \overset{d}{=} \underline{-X}$. (Note: $\overset{d}{=}$ means distributed equally.)

2. One can show $\phi_{\underline{X}}$ is continuous.
3. If A is an $d \times d$ matrix and $\underline{b} \in R^d$, then for random vector $\underline{X} = (X_1, \ldots, X_d)$ we have

$$\begin{aligned}\phi_{A\underline{X}+\underline{b}}(\underline{t}) &= E(e^{i\underline{t}^T(A\underline{X}+\underline{b})}) \\ &= e^{i\underline{t}^T\underline{b}}E(e^{i\underline{t}^TA\underline{X}}) \\ &= e^{i\underline{t}^T\underline{b}}\phi_{\underline{X}}(\underline{t}^TA)\end{aligned}$$

4. If $X_1, \ldots, X_d$ are independent, then

$$\phi_{X_1+\cdots+X_d}(t) = E(e^{it^T\sum_{i=1}^d X_i}) = E(\prod_{j=1}^d e^{it^TX_j}) = \prod_{j=1}^d E(e^{it^TX_j}) = \prod_{j=1}^d \phi_{X_j}(t)$$

Theorem 9.9

1. Uniqueness: $\phi_{\underline{X}}(\underline{t}) = \phi_{\underline{Y}}(\underline{t})$ for all $\underline{t} \in R^d$ iff $\underline{X} \overset{d}{=} \underline{Y}$.
2. Continuity:
   - $X_n \underset{(n\to\infty)}{\overset{d}{\to}} X \Rightarrow \phi_{X_n}(t) \to \phi_X(t)$ for all $t \in R$.
   - If pointwise for all $t \in R$ $\phi(t) := \lim_{n\to\infty} \phi_{X_n}(t)$ exists and is continuous at 0 then $X_n \underset{(n\to\infty)}{\overset{d}{\to}} X$ for a random variable X, with cf ϕ.

9.3.2 Main result

We can show the following:

Lemma 9.11

1. If $a_n \to a$ as $n \to \infty$ then $(1 + \frac{a_n}{n})^n = e^a$ as $n \to \infty$.
2. If $E(|X|^m) < \infty$ for some $m \in N$, then as $t \to 0$,

$$\phi_X(t) = \sum_{k=0}^{m} \frac{(it)^k}{k!} E(X^k) + o(|t|^m)$$

Note: $h(t) \in o(g(t))$ as $t \to 0$ means $\frac{|h(t)|}{|g(t)|} \to 0$ as $t \to 0$.

Theorem 9.12: Central Limit Theorem

If $\{X_n\}_{n \in N}$ is a sequence of iid random variables with $\mu_1 = E(X_1)$ and $\sigma^2 = Var(X_1) < \infty$ then

$$\sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \underset{(n \to \infty)}{\overset{d}{\to}} N(0,1)$$

Remark 9.13

1. For iid random variables with finite second moments, CLT implies $\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$ for large n. Or, $\sum_{j=1}^n X_j \sim N(n\mu, n\sigma^2)$ for large n. If the distribution of X_1 is very different from $N(\mu, \sigma^2)$, a large n should be chosen.
2. If, additionally, $E(|X_1|^3) < \infty$, the Berry-Esseen theorem states the existence of $c \in (\frac{1}{\sqrt{2\pi}}, \frac{1}{2})$ such that $\sup_{x \in R} |F_{\sqrt{n}\frac{\overline{X_n}-\mu}{\sigma}}(x) - \Phi(x)| \leq c\frac{E(|\frac{X_1-\mu}{\sigma}|^3)}{\sqrt{n}}$ for all $n \in N$.
3. Let $\alpha \in (0,1), q_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$. Suppose μ, σ are known. Then:

$$\begin{aligned}
&P(\overline{X_n} \in [\mu - q_\alpha \frac{\sigma}{\sqrt{n}}, \mu + q_\alpha \frac{\sigma}{\sqrt{n}}]) \\
&= P(-q_\alpha \leq \sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} \leq q_\alpha) \\
&= \Phi(q_\alpha) - \Phi(-q_\alpha) = 2\Phi(q_\alpha) - 1 = 2\Phi(\Phi^{-1}(1 - \frac{\alpha}{2})) - 1 = 1 - \alpha
\end{aligned}$$

   Suppose σ is known and μ is not known. Switch the roles of $\overline{X_n}, \mu$. We obtain that: $[\mu - q_\alpha \frac{\sigma}{\sqrt{n}}, \mu + q_\alpha \frac{\sigma}{\sqrt{n}}]$ is (asymptotially for large n) a random interval which contains μ with probability $1 - \alpha$. A so-called asymptotic $(1 - \alpha)$-confidence interval for μ.

Remark 9.15

If X_1 is discrete (with support $S \subseteq Z$), and n is small, one often applies a continuity correction, by computing $P(a - c \leq \sum_{i=1}^n X_i \leq b + c)$ instead of $P(a \leq \sum_{i=1}^n X_i \leq b)$ e.g. for $c = \frac{1}{2}$. If $a = b$, then this is necessary for all n.