MATH50013 - Probability and Statistics for JMC

Notes by Robert Weingart

Contents

1 Introduction

1.1 Introduction to Uncertainty

1.2 Introduction to Statistics

1.2.1 Population vs. Sample

1.3 Probability AND Statistics

1.4 Statistical Modelling

2 Set Theory Review

2.1 Sets, subsets and complements

2.1.1 Sets

2.1.2 Membership, subsets, equality, complements, and singletons

2.2 Set operations

2.2.1 Venn diagrams, Unions and Intersections

2.2.2 Cartesian Products

2.3 Cardinality

3 Visual and Numerical Summaries

3.1 Visualization

3.1.1 The histogram

Definition. A **histogram** partitions the range of a sample into **bins** and shows what number of data points in each bin. Rather than frequency, the amount shown can also be relative frequency or density.

3.1.2 Empirical CDF

Definition. The **indicator function** is defined as $I(\text{false}) := 0$ and $I(\text{true}) = 1$.

Definition. The **empirical cumulative distribution function** of a sample is

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} I(x_i \leq x)$$

3.2 Summary Statistics

3.2.1 Measures of Location

Definition. The **arithmetic mean** is $\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$.

Definition. The **geometric mean** is $x_G := \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$.

Definition. The **harmonic mean** is $x_H := n\left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^{-1}$

Definition. The i**th order statistic**, written $x_{(i)}$, is the ith smallest value of the sample. For non-integer values of the form $i + \alpha$ with $\alpha \in (0, 1)$, we define

$$x_{(i+\alpha)} := (1 - \alpha)x_{(i)} + \alpha x_{(i+1)}$$

Definition. The **median** is $x_{(\frac{n+1}{2})}$.

Definition. The **mode** is the most frequently occurring value. If there are multiple then the sample is **multimodal**.

3.2.2 Measures of Dispersion

Definition. The **mean square** or **sample variance** is

$$s_x^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Definition. The **root mean square** or **sample standard deviation** is

$$s_x := \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Definition. The **range** is $x_{(n)} - x_{(1)}$.

Definition. The **first quartile** is $x_{\left(\frac{1}{4}(n+1)\right)}$. The **third quartile** is $x_{\left(\frac{3}{4}(n+1)\right)}$. The **interquartile range** is the difference between the third and first quartiles.

3.2.3 Covariance and Correlation

Definition. For a sample where each data point is an (x_i, y_i) pair, the **covariance** is

$$s_{xy} := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\bar{y}$$

.

Definition. For a sample as above, the **correlation** is

$$r_{xy} := \frac{s_{xy}}{s_x s_y}$$

3.2.4 Skewness

Definition. The **skewness** is $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$.

3.3 One more visualization: the box-and-whisker plot

Definition. A **box-and-whisker plot** shows the median, first and third quartiles, points within $\frac{3}{2} \times IQR$ of the quartiles, and any outliers.

4 Probability

4.1 The formal structure

4.1.1 σ-algebras

Definition 4.1.1. A **σ-algebra associated with** S is a set $\mathcal{F}$ of subsets of S where $S \in \mathcal{F}$, $\mathcal{F}$ is closed under complements with respect to S, and $\mathcal{F}$ is closed under countable unions.

Proposition. $\emptyset \in \mathcal{F}$. *$\mathcal{F}$ is also closed under countable intersections.*

4.1.2 Probability measure

Definition 4.1.2. A **probability measure** is a function $P : \mathcal{F} \to \mathbb{R}$ where $P(E) \geq 0$ for any E, $P(S) = 1$, and for countably many disjoint sets E_i,

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

. A triple $(S, \mathcal{F}, P)$ as previously defined is a **probability space**.

4.2 Interpretations of the probability space

4.3 Interpretation of the σ-algebra

4.3.1 The sample space (S)

Definition. The **sample space** S is the set of all possible outcomes of an experiment.

4.3.2 The event space ($\mathcal{F}$)

Definition. An **event** is a subset $E \subset S$. $\mathcal{F}$ is the set of all possible events being considered (which may not include all possible combinations of outcomes).

Definition. E_1 and E_2 are **mutually exclusive** iff $E_1 \cap E_2 = \emptyset$ i.e. they cannot both happen at once.

4.4 Interpretations of the probability measure (P)

4.4.1 Classical interpretation

Definition. In the **classical interpretiation**, S consists of finitely many equally likely **elementary events** and $P(E) = \frac{|E|}{|S|}$. For an infinite S, this can still be applied by replacing cardinality above with a different measure.

4.4.2 Frequentist interpretation

Definition. In the **frequentist interpretation**, when an experiment is repeated infinitely many times, the proportion of trials in which E occurs approaches $P(E)$.

4.4.3 Subjective interpretation

Definition. In the **subjective interpretation**, $P(E)$ is the degree of belief a person has that E occurs.

4.5 A few derivations from the axioms

Proposition. *For $E, F \in \mathcal{F}$,*

- $P(\emptyset) = 0$
- $P(E) \leq 1$
- $P(\overline{E}) = 1 - P(E)$
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- $P(E \cap \overline{F}) = P(E) - P(E \cap F)$
- $E \subseteq F \implies P(E) \leq P(F)$

4.6 Conditional Probability

Definition 4.6.1. For $P(F) > 0$ the **conditional probability of E given F** is

$$P(E \mid F) := \frac{P(E \cap F)}{P(F)}$$

Proposition. *For $P(F) > 0$,*

- *For any $E \in \mathcal{F}$, $P(E \mid F) \geq 0$*
- $P(F \mid F) = 1$
- *For $E_1, \dots, E_n \in \mathcal{F}$ pairwise disjoint, $P\left(\bigcup_{i=1}^n E_i \mid F\right) = \sum_{i=1}^n P(E_i \mid F)$*

4.7 Independent Events

Definition 4.7.1. $E, F \in \mathcal{F}$ are **independent** iff $P(E \cap F) = P(E)P(F)$. $E_1, \dots E_n$ are **independent** iff for any subset $E_{i_1}, \dots, E_{i_l}$ we have $P\left(\bigcap_{j=1}^l E_{i_j}\right) = \prod_{j=1}^l P(E_{i_j})$.

Proposition. *E and F are independent $\implies$ E and $\overline{F}$ are independent.*

Proposition. *E and F are independent $\iff$ $P(E \mid F) = P(E)$.*

4.7.1 More Examples

4.7.2 Conditional Independence

Definition. For $E_1, E_2, F \in \mathcal{F}$, E_1 and E_2 are **conditionally independent given** F iff $P(E_1 \cap E_2 \cap F) = P(E_1 \mid F)P(E_2 \mid F)$.

4.7.3 Joint Events

Definition. When combining multiple independent experiments, a **probability table** can be used to show the probabilities of all elementary events (i.e. combinations of an elementary event in each experiment).

4.8 Bayes's Theorem

Theorem 4.9. *(Bayes's) For $E, F \in \mathcal{F}$ with $P(E) > 0$ and $P(F) > 0$,*

$$P(E \mid F) = \frac{P(F \mid E)P(E)}{P(F)}$$

Theorem 4.10. *(The Law of Total Probability) For a partition $E_1, \dots$ of S, and any $F \in \mathcal{F}$, $P(F) = \sum_i P(F \mid E_i)P(E_i)$.*

Theorem 4.11. *(Bayes's applied to a partition) For a partition $E_1, \dots$ of S with $P(E_i) > 0$ for all i and $F \in \mathcal{F}$ with $P(F) > 0$,*

$$P(E_i \mid F) = \frac{P(F \mid E_i)P(E_i)}{\sum_j P(F \mid E_j)P(E_j)}$$

4.12 More Examples

5 Discrete Random Variables

5.1 Random Variables

Definition 5.1.1. A **random variable** is a measurable mapping $X : S \to \mathbb{R}$ where $\forall x \in \mathbb{R},\ \{s \in S : X(s) \leq x\} \in \mathcal{F}$.

Definition 5.1.2. The **range** of X is $\mathbb{X}$, the image of S under X.

Definition. The **probability distribution** of X is

$$P_X(X \in B) := P(\{s \in S : X(S) \in B\})$$

where $B \subseteq \mathbb{R}$.

Notation. For brevity we write $\{X \in B\} := \{s \in S : X(s) \in B\}$ (TODO: doesn't this make P and P_X interchangeable?) and $\{a < X \leq b\} := \{X \in (a, b]\}$ etc.

5.1.1 Cumulative Distribution Function

Definition 5.1.3. The **cumulative distribution function** of X is $F_X : \mathbb{R} \to [0, 1]$ where $F_X(x) = P_X(X \leq x)$.

Definition. A function f is **right-continuous** iff for any decreasing sequence $x_i \to x$ we have $f(x_i) \to f(x)$.

Proposition. *A CDF is right-continuous.*

Proposition. *F_X is a CDF iff all the following hold:*

- *F_X is right-continuous*
- $F_X(\mathbb{R}) \subseteq [0, 1]$
- *F_X is monotonically increasing*
- $\lim_{x\to-\infty} F_X(x) = 0$
- $\lim_{x\to\infty} F_X(x) = 1$

5.2 Discrete Random Variables

Definition 5.2.1. A random variable is **discrete** iff its range is finite or countably infinite.

Definition 5.2.2. For a DRV X, the **probability mass function** $p_X : \mathbb{R} \to [0, 1]$ is $p_X(x) = P_X(X = x)$ for $x \in \mathbb{X}$ and $p_X(x) = 0$ for $x \notin \mathbb{X}$.

Definition. The **support** of X is $\{x \in \mathbb{R} : p_X(x) > 0\}$. Usually this is $\mathbb{X}$.

5.2.1 Properties of Mass Function p_X

Proposition. *An arbitrary function p_X can be a PMF for X iff $\forall x \in \mathbb{X},\ p_X(x) \geq 0$ and $\sum_{x\in\mathbb{X}} p_X(x) = 1$.*

5.2.2 Discrete Cumulative Distribution Function

Definition. The **cumulative distribution function** of a DRV X is $F_X(x) = P(X \leq x)$ (TODO: is this not what it always is?).

5.2.3 Connection between F_X and p_X

Proposition. *For $\mathbb{X} = \{x_1, \ldots\}$ with the $x_i \leq x_{i+1}$ for all i,*

$$F_X(x) = \sum_{x_i \leq x} p_X(x_i)$$

Equivalently,

$$\forall i \geq 1,\ p_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

5.2.4 Properties of Discrete CDF F_X

Proposition. *We have*

- $\lim_{x\to-\infty} F_X(x) = 0$
- $\lim_{x\to\infty} F_X(x) = 1$
- $\lim_{h\to 0^+} F_X(x+h) = F_X(x)$
- $a < b \implies F_X(a) \leq F_X(b)$
- *For* $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a)$

5.3 Functions of a discrete random variable

Proposition. *For a DRV X and $g : \mathbb{X} \to \mathbb{R}$, $Y = g(X)$ is also a DRV. We have*

$$p_Y(y) = \sum_{x\in\mathbb{X}:g(x)=y} p_X(x)$$

5.4 Mean and Variance

5.4.1 Expectation

Definition 5.4.1. The **expected value** or **mean** of a DRV X is

$$E_X(X) := \sum_{x\in\mathbb{X}} x p_X(x)$$

It is often abbreviated to $E(X)$.

Theorem 5.5. *For a* ***function of interest*** $g : \mathbb{R} \to \mathbb{R}$, *we have*

$$E(g(X)) = \sum_{x\in\mathbb{X}} g(x) p_X(x)$$

Proposition. *E is linear.*

Definition 5.5.1. For a DRV X, the **variance** of X is

$$\mathrm{Var}_X(X) := E_X\left((X - E_X(X))^2\right) = E(X^2) - E(X)^2$$

Proposition. *For $a, b \in \mathbb{R}$, $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$*

Definition 5.5.2. For a DRV X, the **standard deviation** of X is

$$\mathrm{sd}(X) := \sqrt{\mathrm{Var}_X(X)}$$

Definition 5.5.3. For a DRV X, the **skewness** of X is

$$\gamma_1 := \frac{E_X((X - E_X(X))^3)}{\mathrm{sd}_X(X)^3}$$

5.5.1 Sums of Random Variables

Proposition. *For $X_1, \dots X_n$ (possibly with different distributions, not necessarily independent) with sum S_n, we have*

$$E(S_n) = \sum_{i=1}^{n} E(X_i)$$

and

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

Proposition. *For $X_1, \dots X_n$ independent with sum S_n, we have*

$$\operatorname{Var}(S_n) = \sum_{i=1}^{n} \operatorname{Var}(X_i)$$

and

$$\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}(X_i)$$

Proposition. *For $X_1, \dots X_n$ independent and identically distributed with sum S_n, $E(X_i) = \mu$ and $\operatorname{Var}(X_i) = \sigma^2$, we have*

$$E\left(\frac{S_n}{n}\right) = \mu$$

and $\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$

5.6 Some Important Discrete Random Variables

X	$\mathbb{X}$	$p_X(x)$	$E(X)$	$\operatorname{Var}(X)$	γ_1
$X \sim \text{Bernoulli}(p)$	$\{0, 1\}$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$\frac{1-2p}{\sqrt{p(1-p)}}*$
$X \sim \text{Binomial}(n, p)$	$\{0, \dots n\}$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$
$X \sim \text{Geometric}(p)$	$\{1, 2, \dots\}$	$p(1-p)^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{2-p}{\sqrt{1-p}}$
$X \sim \text{Poisson}(\lambda)$	$\{0, 1, \dots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	$\frac{1}{\sqrt{\lambda}}$
$X \sim \text{U}(\{1, \dots, n\})$	$\{1, \dots, n\}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	0

$*$: The skewness of the Bernoulli distribution is not given in the official notes.

5.6.1 Bernoulli Distribution

$X \sim \text{Bernoulli}(p)$ chooses between 1 and 0 where $P(X = 1) = p$.

5.6.2 Binomial Distribution

$X \sim \text{Binomial}(n, p)$ is the total number of successes after n Bernoulli trials with probability p.

5.6.3 Geometric Distribution

$X \sim \text{Geometric}(p)$ is the number of Bernoulli trials with probability p it will take to have the first success.

5.6.4 Poisson Distribution

$X \sim \text{Poisson}(\lambda)$ is the number of occurrences of an event that occurs at a rate of λ.

5.6.5 Discrete Uniform Distribution

$X \sim \text{U}(\{1, \ldots, n\})$ is a random value out of $\{1, \ldots n\}$.

6 Continuous Random Variables

Definition 6.0.1. A random variable X is absolutely **continuous** iff there exists a measurable non-negative function $f_X : \mathbb{R} \to \mathbb{R}$ (the **probability density function**) where

$$\forall B \subseteq \mathbb{R},\ P(X \in B) = \int_{x \in B} f_X(x) dx$$

6.0.1 Continuous Cumulative Distribution Function

Definition 6.0.2. The **cumulative distribution function** of a CRV X is $F_X(x) = P(X \leq x)$ (as for any RV).

Proposition. *For a CRV X, $F_X(x) = \int_{-\infty}^{x} f_X(x')dx'$*

6.0.2 Properties of Continuous F_X and f_X

Proposition. *For a CRV X,*

- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$
- *If F_X is differentiable at x then $f_X(x) = F'_X(x)$*
- $\forall a \in \mathbb{R},\ P(X = a) = 0$
- *For $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a)$*
- *$f_X(X)$ is not a probability, so we do not require $f_X(x) \leq 1$*
- *X is uniquely defined by f_X*

Proposition. *An arbitrary function f_X is a PDF for a CRV iff $\forall x \in \mathbb{R},\ f_X(x) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x)dx = 1$ (f_X is **normalised**).*

6.0.3 Transformations

Proposition. *For $Y = g(X)$ with g strictly monotonically increasing, we have*

$$F_Y(y) = F_X(g^{-1}(y))$$

and

$$f_Y(y) = f_X\left(g^{-1}(y)\right) g^{-1'}(y)$$

Proposition. *For $Y = g(X)$ with g strictly monotonically decreasing, we have*

$$F_Y(y) = 1 - F_X(g^{-1}(y))$$

and

$$f_Y(y) = -f_X\left(g^{-1}(y)\right) g^{-1'}(y)$$

6.1 Mean, Variance and Quantiles

6.1.1 Expectation

Definition 6.1.1. The **mean** or **expectation** of a CRV X is

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

Definition. For any measurable **function of interest** $g : \mathbb{R} \to \mathbb{R}$ we have

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

Proposition. *E is linear.*

6.1.2 Variance

Definition 6.1.2. The **variance** of a CRV X is

$$\mathrm{Var}_X(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$$

Proposition. *For $a, b \in \mathbb{R}$, $\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)$*

6.1.3 Quantiles

Definition 6.1.3. For $\alpha \in [0, 1]$, we α-**quantile** of a CRV X is

$$Q_X(\alpha) := F_X^{-1}(\alpha)$$

so that $P(X \le Q_X(\alpha)) = \alpha$.

6.2 Some Important Continuous Random Variables

X	$\mathbb{X}$	$f_X(x)$	$F_X(x)$	$E(X)$	$\mathrm{Var}(X)$
$X \sim \mathrm{U}(a, b)$	(a, b)	$\begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$X \sim \mathrm{Exp}(\lambda)$	$[0, \infty)$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$X \sim \mathbb{N}(\mu, \sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$	μ	σ^2

6.2.1 Continuous Uniform Distribution

$X \sim \mathrm{U}(a,b)$ or $X \sim \mathrm{Uniform}(a,b)$ is uniformly distributed on the interval (a,b) and 0 elsewhere.

Definition. Tht **standard uniform** is $\mathrm{Uniform}(0,1)$.

Proposition. $X \sim \mathrm{Uniform}(0,1) \implies (a+(b-a)X) \sim \mathrm{Uniform}(a,b)$.

6.2.2 Exponential Distribution

$X \sim \mathrm{Exp}(\lambda)$ is the time until an event occurring at rate λ occurs.

Proposition. *$X \sim \mathrm{Exp}(\lambda)$ exhibits the **Lack of Memory Property**:*

$$\forall x, t > 0,\ P(X > t + x \mid X > t) = P(X > x)$$

Proposition. *If the number of events occurring in an interval of size x is $N_x \sim \mathrm{Poisson}(\lambda x)$ then the separation between two events is $X \sim \mathrm{Exp}(\lambda)$.*

6.2.3 Normal (Gaussian) Distribution

$X \sim N(\mu, \sigma^2)$ has no obvious interpretation.

Definition. $X \sim N(0,1)$ is the **standard normal distribution** or **unit normal distribution**. It has the PDF

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

and the CDF

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

Proposition. $X \sim N(0,1) \implies (\sigma X + \mu) \sim N(\mu, \sigma^2)$

Theorem 6.3. *(Central Limit Theorem) For $X_1, \ldots, X_n$ independent and identically distributed with mean μ and variance σ^2,*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

6.4 Further examples

7 Joint Random Variables

8 Estimation

9 Hypothesis Testing

10 Convergence Concepts