



THE UNIVERSITY OF QUEENSLAND
A U S T R A L I A

COURSE NOTES FOR STAT3001
MATHEMATICAL STATISTICS

CONTRIBUTORS:

MICHAEL CICCOTOSTO-CAMP
NAME2

THE UNIVERSITY OF QUEENSLAND
SCHOOL OF MATHEMATICS AND PHYSICS

CONTENTS

SYMBOLS AND NOTATION	iii
REVIEW	1
USEFUL FORMULAE AND THEOREMS	1
COMMON DISTRIBUTIONS	2
COMMON PROBABILISTIC PROPERTIES AND IDENTITIES	3
PROBABILISTIC PROPERTIES	3
PROBABILISTIC IDENTITIES	4
POINT ESTIMATION	5
METHODS OF FINDING ESTIMATES	5
REFERENCES	6

SYMBOLS AND NOTATION

Matrices are capitalized bold face letters while vectors are lowercase bold face letters.

<i>Syntax</i>	<i>Meaning</i>
\triangleq	An equality which acts as a statement
$ \mathbf{A} $	The determinate of a matrix.
$\mathbf{x}^\top, \mathbf{X}^\top$	The transpose operator.
$\mathbf{x}^*, \mathbf{X}^*$	The hermitian operator.
$\mathbf{a}.*\mathbf{b}$ or $\mathbf{A}.*\mathbf{B}$	Element-wise vector (matrix) multiplication, similar to Matlab.
\propto	Proportional to.
∇ or ∇_f	The partial derivative (with respect to f).
$\nabla\nabla$ or $H(f)$	The Hessian.
\sim	Distributed according to, example $X \sim \mathcal{N}(0, 1)$
$\overset{\text{iid}}{\sim}$	Identically and independently distributed according to, example $X_1, X_2, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$
$\mathbf{0}$ or $\mathbf{0}_n$ or $\mathbf{0}_{n \times m}$	The zero vector (matrix) of appropriate length (size) or the zero vector of length n or the zero matrix with dimensions $n \times m$.
$\mathbf{1}$ or $\mathbf{1}_n$ or $\mathbf{1}_{n \times m}$	The one vector (matrix) of appropriate length (size) or the one vector of length n or the one matrix with dimensions $n \times m$.
$\mathbb{1}_{n \times m}$	The matrix with ones along the diagonal and zeros on off diagonal elements.

$\mathbf{A}_{(:,)}$	Index slicing to extract a submatrix from the elements of $\mathbf{A} \in \mathbb{R}^{n \times m}$, similar to indexing slicing from the python and Matlab programming languages. Each parameter can receive a single value or a 'slice' consisting of a start and an end value separated by a semicolon. The first and second parameter describe what row and columns should be selected, respectively. A single value means that only values from the single specified row/column should be selected. A slice tells us that all rows/columns between the provided range should be selected. Additionally if now start and end values are specified in the slice then all rows/columns should be selected. For example, the slice $\mathbf{A}_{(1:3,j:j')}$ is the submatrix $\mathbb{R}^{3 \times (j'-j+1)}$ matrix containing the first three rows of \mathbf{A} and columns j to j' . As another example, $\mathbf{A}_{(:,j)}$ is the j^{th} column of \mathbf{A} .
\mathbf{A}^\dagger	Denotes the unique psuedo inverse or Moore-Penore inverse of \mathbf{A} .
\mathbb{C}	The complex numbers.
$\text{diag}(\mathbf{w})$	Vector argument, a diagonal matrix containing the elements of vector \mathbf{w} .
$\text{diag}(\mathbf{W})$	Matrix argument, a vector containing the diagonal elements of the matrix \mathbf{W} .
\mathbb{E} or $\mathbb{E}_{q(x)}[z(x)]$	Expectation, or expectation of $z(x)$ where $x \sim q(x)$.
\mathbb{R}	The real numbers.
$\text{tr}(\mathbf{A})$	The trace of a matrix.
\mathbb{V} or $\mathbb{V}_{q(x)}[z(x)]$	Variance, the variance of $z(x)$ when $x \sim q(x)$.
\mathbb{Z}	The integers, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$.
Ω	The sample space.

REVIEW

Theorems and definitions here are mostly concepts seen before from other courses.

Useful Formulae and Theorems.

(Geometric Series)
$$\sum_{k=0}^{n-1} r^k = \left(\frac{1 - r^n}{1 - r} \right)$$

or

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1 - r} \quad \text{with} \quad |r| < 1$$

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

(Newton's Binomial formula)
$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

Theorem 1 (Young's inequality for products). *If $a \geq 0$ and $b \geq 0$ are nonnegative real numbers and if $p > 1$ and $q > 1$ are real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Equality holds iff $a^p = b^q$.

Common Distributions. Common distributions seen from prior courses. Notations mostly borrowed from STAT2003.

<i>Name</i>	<i>Notation</i>	<i>Support</i>	<i>pf</i>	<i>Expectation</i>	<i>Variance</i>
Bernoulli	$\text{Ber}(p)$	$\{0, 1\}$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Binomial	$\text{Bin}(n, p)$	$\{0, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Negative-Binomial	$\text{NB}(r, p)$	\mathbb{N}_0	$\binom{x+r-1}{x} p^x (1-p)^r$	$\frac{rp}{1-p}$	$\frac{rp}{(1-p)^2}$
Geometric	$\text{Geo}(n, p)$	\mathbb{N}_0	$(1-p)^k p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson	$\text{Poi}(\lambda)$	\mathbb{N}_0	$\frac{\lambda^x}{x!} e^{-\lambda}$	λ	λ
Uniform	$\text{U}[a, b]$	$[a, b]$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Exponential	$\text{Exp}(\lambda)$	\mathbb{R}^+	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda}$
Normal	$\text{N}(\mu, \sigma^2)$	\mathbb{R}	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	μ	σ^2
Gamma	$\text{Gam}(\alpha, \lambda)$	\mathbb{R}^+	$\frac{\lambda^\alpha x^{\alpha-1} \exp(-\lambda x)}{\Gamma(\alpha)}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Chi-Squared	χ_n^2	\mathbb{R}^+	$\frac{x^{\frac{n}{2}-1} \exp(-\frac{1}{2}x)}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}$	n	$2n$
White-Noise	$\text{WN}(\mu, \sigma^2)$	NA	NA	μ	σ^2

Common Probabilistic Properties and Identities. Common probabilistic properties seen from prior courses.

Probabilistic Properties. For any random variables, the following hold.

$$(1) \quad \mathbb{E}(X) = \int_0^\infty (1 - F(X)) \, dx$$

$$(2) \quad \mathbb{E}(aX + b) = a\mathbb{E}X + b$$

$$(3) \quad \mathbb{E}(g(X) + h(X)) = \mathbb{E}g(X) + \mathbb{E}h(X)$$

$$(4) \quad \text{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

$$(5) \quad \text{Var}(aX + b) = a^2\text{Var}(X)$$

$$(6) \quad \text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$$

$$(7) \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$(8) \quad \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]]$$

$$(9) \quad \text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X])$$

$$(10) \quad |\mathbb{E}(XY)|^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

$$(11) \quad |\text{Cov}(XY)|^2 \leq \text{Var}(X)\text{Var}(Y)$$

$$(12) \quad \mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

$$(\text{Bayes' Theorem}) \quad \mathbb{P}(A | B) = \frac{\mathbb{P}(B | A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

$$(13) \quad \mathbb{P}(A_1, \dots, A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1, A_2) \cdots \mathbb{P}(A_n | A_1, A_2, \dots, A_{n-1})$$

$$(14)$$

Let $\Omega = \bigcup_{i=1}^n B_i$ (that is B_i partitions the sample space) then

$$(\text{TLoP}) \quad \mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A | B_i)\mathbb{P}(B_i)$$

$$(\text{TLoE}) \quad \mathbb{E}(A) = \sum_{i=1}^n \mathbb{E}(A | B_i)\mathbb{P}(B_i)$$

which, when **TLoP** used in conjunction with Bayes' Rule gives

$$(15) \quad \mathbb{P}(B_i | A) = \frac{\mathbb{P}(A | B_i)\mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A | B_j)\mathbb{P}(B_j)}.$$

If $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \text{WN}(\mu, \sigma^2)$ and $S_n = \sum_{i=1}^n X_i$, then for all $\epsilon > 0$

$$(\text{Weak Law of Large Numbers}) \quad \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0.$$

If $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \text{WN}(\mu, \sigma^2)$ and $S_n = \sum_{i=1}^n X_i$, then for all $x \in \mathbb{R}$

$$(CLT) \quad \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x).$$

If X is a random variable and h is a convex function then

$$(\text{Jensens Inequality}) \quad h(\mathbb{E}(X)) \leq \mathbb{E}(h(X)).$$

Probabilistic Identities. If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Ber}(p)$ then

$$(16) \quad \sum_{i=1}^n X_i \sim \text{Bin}(n, p).$$

If $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, then $X + Y \sim \text{Bin}(n + m, p)$.

If $X \sim \text{N}(\mu_X, \sigma_X^2)$ and $Y \sim \text{N}(\mu_Y, \sigma_Y^2)$, then $X + Y \sim \text{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

If $X_1, X_2, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma)$ then

$$(17) \quad \sum_{i=1}^n X_i^2 = \chi_n^2.$$

POINT ESTIMATION

Methods of Finding Estimates.

Definition 2 (Statistic). Let X_1, \dots, X_n be a random sample of size n from a population and let $T(x_1, \dots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of (X_1, \dots, X_n) . The random variable or random vector $Y = T(X_1, \dots, X_n)$ is called a **statistic**. The probability distribution of a statistic Y is called the **sampling distribution** of Y [Cas01, page 211].

Definition 3 (Sample Mean). The **sample mean** is the arithmetic average of the values in a random sample. It is usually denoted by

$$(18) \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

[Cas01, page 212].

Definition 4 (Sample Variance and Standard Deviation). The **sample variance** is the statistic defined by

$$(19) \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

The **sample standard deviation** is the statistic defined by $S = \sqrt{S^2}$ [Cas01, page 212].

Definition 5 (Sufficient Statistic). A statistic $T(\mathbf{X})$ is a **sufficient statistic** for θ if the conditional distribution of the sample \mathbf{X} given the value of $T(\mathbf{X})$ does not depend on θ [Cas01, page 272].

REFERENCES

[Cas01] George and Berger Casella Roger, *Statistical Inference*, Cengage, Mason, OH, 2001 (eng).