THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

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 v
    Useful Formulae and Theorems v
    Common Distributions vi
    Common Probabilistic Properties and Identities vii
        Probabilistic Properties vii
        Probabilistic Identities 1
References 2

Symbols and Notation

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

Syntax	*Meaning*
$\triangleq$	An equality which acts as a statement
$\lvert \boldsymbol{A} \rvert$	The determinate of a matrix.
$\boldsymbol{x}^{\intercal}, \boldsymbol{X}^{\intercal}$	The transpose operator.
$\boldsymbol{x}^{*}, \boldsymbol{X}^{*}$	The hermitian operator.
$\boldsymbol{a}. * \boldsymbol{b}$ or $\boldsymbol{A}. * \boldsymbol{B}$	Element-wise vector (matrix) multiplication, similar to Matlab.
$\propto$	Proportional to.
∇ or $\nabla_{\boldsymbol{f}}$	The partial derivative (with respect to $\boldsymbol{f}$).
$\nabla\nabla$ 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 $X_1, X_2, \ldots X_n \sim \mathcal{N}(0,1)$
$\mathbf{0}$ 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.

$\boldsymbol{A}_{(\cdot,\cdot)}$	Index slicing to extract a submatrix from the elements of $\boldsymbol{A} \in \mathbb{R}^{n\times m}$, 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 $\boldsymbol{A}_{(1:3,j:j')}$ is the submatrix $\mathbb{R}^{3\times(j'-j+1)}$ matrix containing the first three rows of $\boldsymbol{A}$ and columns j to j'. As another example, $\boldsymbol{A}_{(:,j)}$ is the j^{th} column of $\boldsymbol{A}$.
$\boldsymbol{A}^{\dagger}$	Denotes the unique psuedo inverse or Moore-Penore inverse of $\boldsymbol{A}$.
$\mathbb{C}$	The complex numbers.
$\mathrm{diag}\left(\boldsymbol{w}\right)$	Vector argument, a diagonal matrix containing the elements of vector $\boldsymbol{w}$.
$\mathrm{diag}\left(\boldsymbol{W}\right)$	Matrix argument, a vector containing the diagonal elements of the matrix $\boldsymbol{W}$.
$\mathbb{E}$ or $\mathbb{E}_{q(x)}\left[z(x)\right]$	Expectation, or expectation of $z(x)$ where $x \sim q(x)$.
$\mathbb{R}$	The real numbers.
$\mathrm{tr}\left(\boldsymbol{A}\right)$	The trace of a matrix.
$\mathbb{V}$ or $\mathbb{V}_{q(x)}\left[z(x)\right]$	Variance, the variance of $z(x)$ when $x \sim q(x)$.
$\mathbb{Z}$	The integers, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
Ω	The sample space.

Review

Theorems and defintions 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 $\frac{1}{p} + \frac{1}{p} = 1$, 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.

Name	*Notation*	*Support*	*pf*	*Expectation*	*Variance*
Bernoulli	$\mathsf{Ber}(p)$	$\{0,1\}$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Binomial	$\mathsf{Bin}(n,p)$	$\{0,\ldots,n\}$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Geometric	$\mathsf{Geo}(n,p)$	$\mathbb{N}_0$	$(1-p)^k p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson	$\mathsf{Poi}(\lambda)$	$\mathbb{N}_0$	$\frac{\lambda^x}{x!}e^{-\lambda}$	λ	λ
Uniform	$\mathsf{U}[a,b]$	$[a,b]$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Exponential	$\mathsf{Exp}(\lambda)$	$\mathbb{R}^+$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda}$
Normal	$\mathsf{N}(\mu,\sigma^2)$	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	μ	σ^2
Gamma	$\mathsf{Gam}(\alpha,\lambda)$	$\mathbb{R}^+$	$\frac{\lambda^\alpha x^{\alpha-1}\exp(-\lambda x)}{\Gamma(\alpha)}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Chi-Squared	χ^2_n	$\mathbb{R}^+$	$\frac{x^{\frac{n}{2}-1}\exp(-\frac{1}{2}x)}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}$	n	$2n$
White-Noise	$\mathsf{WN}(\mu,\sigma^2)$	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.

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

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

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

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

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

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

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

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] \tag{8}$$

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

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

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

$$\mathbb{P}\,(A \mid B) = \frac{\mathbb{P}\,(A \cap B)}{\mathbb{P}\,(B)} \tag{12}$$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{Bayes' Theorem}$$

$$\mathbb{P}\,(A_1, \ldots, A_n) = \mathbb{P}\,(A_1)\,\mathbb{P}\,(A_2 \mid A_1)\,\mathbb{P}\,(A_3 \mid A_1, A_2) \cdots \mathbb{P}\,(A_n \mid A_1, A_2, \ldots, A_{n-1}) \tag{13}$$

(14)

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

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i) \tag{TLoP}$$

$$\mathbb{E}(A) = \sum_{i=1}^{n} \mathbb{E}(A \mid B_i)\mathbb{P}(B_i) \tag{TLoE}$$

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

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

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

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

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

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

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

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

Probabilistic Identities.

References

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