# Math 541A Midterm 1 Cheat Sheet

Gregory Faletto

# 1 Review of Probability Theory

**Proposition 1 (Change of Variables).** If $U$ is a "nice" subset of $\mathbb{R}^2$ and $\phi$ is an injective differentiable function on $U$, then

$$\int_{\phi(U)} f(u,v)dudv = \int_U f(\phi(x,y))|J\phi(x,y)|dxdy$$

where $J\phi(x,y)$ is the Jacobian of $\phi$ at $(x,y)$.

**Definition 1.1.** Random variables $X_1, X_2, \ldots, X_n$ are independent if for every $B_1, B_2, \ldots, B_n \subseteq \mathbb{R}$, the events $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \ldots, \{X_n \in B_n\}$ are independent; that is,

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

# 2 Limit Theorems

**Definition 2.1. Convergence in probability.** $\{X_n\}$ is said to **converge in probability** to $X$ if

- Grimmett and Strizaker definition:

$$\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0, \text{ for every } \epsilon > 0$$

- More formal (from Math 541A):

$$\forall\epsilon > 0, \lim_{n\to\infty} \Pr(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon) = 0$$

**Definition 2.2. Convergence with probability 1 or almost surely.** The sequence of random variables $\{X_n\}$ is said to **converge with probability 1** (or **almost surely**) to $X$ if

$$\Pr\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$$

**Remark.** This is often written as $X_n \xrightarrow{w.p.1} X$ or $X_n \xrightarrow{a.s.} X$. An equivalent condition for convergence with probability 1 is given by

$$\lim_{n\to\infty} \Pr(|X_m - X| < \epsilon, \text{ for all } m \geq= n) = 1, \text{ for every } \epsilon > 0$$

which shows that convergence in probability is a special case of convergence with probability 1 (obtained by setting $m = n$). Convergence with probability 1 is stronger than convergence in probability and is often referred to as "strong convergence."

**Definition 2.3. Convergence in $r$-th mean** or **convergence in $\ell_p$.** $X_n \to X$ **in $r$th mean** (or **in $\ell_p$**) where $r \geq 1$ (or $0 < p \leq \infty$) if $\mathbb{E}|X_n^r| < \infty$ for all $n$ and

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^r) = 0$$

or if $\|X\|_p < \infty$ and

$$\lim_{n\to\infty} \|X_n - X\|_p = 0$$

**Remark.** Recall that $\|X\|_p := (\mathbb{E}(X)^p)^{1/p}$ if $0 < p < \infty$ and $\|X\|_\infty := \inf\{c > 0 : \Pr(|X| \leq c) = 1\}$. Note that if $p < 1$, $\|\cdot\|_p$ is no longer a norm because it does not satisfy the Triangle Inequality (Corollary **??** and Theorem **??**), but this property still holds. Convergence in $r$th mean is often written $X_n \xrightarrow{r} X$.

**Definition 2.4. Convergence in Distribution.** Let $X_1, X_2, \ldots$ have distribution functions $F_1(\cdot), F_2(\cdot), \ldots$ respectively. Then $X_n$ is said to **converge in distribution to** $X$ if

$$\lim_{n\to\infty} \Pr(X_n \leq u) = \Pr(X \leq u)$$

for all $u$ at which $F_X(x) = \Pr(X \leq x)$ is continuous.

# 3 Exponential Families

**Definition 3.1.** Let $n, k$ be positive integers and let $\mu$ be a *measure* on $\mathbb{R}^n$ (that is, a probability law that does not necessarily sum to 1). Let $t_1, \ldots, t_k : \mathbb{R}^n \to \mathbb{R}$. Let $h : \mathbb{R}^n \to [0, \infty]$, and assume $h$ is not identically zero. For any $w = (w_1, \ldots, w_k) \in \mathbb{R}^k$, define

$$a(w) := \log \left[ \int_{\mathbb{R}^n} h(x) \exp\left( \sum_{i=1}^k w_i t_i(x) \right) d\mu(x) \right], \quad \forall x \in \mathbb{R}^n$$

The set $\{w \in \mathbb{R}^k\}$ is called the **natural parameter space**. On this set, the function

$$f_w(x) := h(x) \exp\left( \sum_{i=1}^k w_i t_i(x) - a(w) \right), \quad \forall x \in \mathbb{R}^n$$

satisfies $\int_{\mathbb{R}^n} f_w(x) d\mu(s) = 1$ (by the definition of $a(w)$). So, the set of functions (which can be interpreted as probability density functions, or as probability mass functions according to $\mu$) $\{f_w : \theta \in \Theta : a(w(\theta)) < \infty\}$ is called a $k$**-parameter exponential family in canonical form.**

More generally, let $\Theta \in \mathbb{R}^k$ be any set an let $w : \Theta \to \mathbb{R}^k$. We define a $k$**-parameter exponential family** to be a set of functions $\{f_\theta : \theta \in \Theta\}$, where

$$f_\theta(x) := h(x) \exp\left( \sum_{i=1}^k w_i(\theta) t_i(x) - a(w(\theta)) \right), \quad \forall x \in \mathbb{R}^n$$

**Theorem 2** (**Theorem 3.4.2 from Casella and Berger**). If $X$ is a random variable in an exponential family, then

$$\mathbb{E}\left( \sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X) \right) = \frac{\partial}{\partial \theta_j} a(w(\theta)). \tag{1}$$

**Proposition 3** (**Way we did this in class**).

$$e^{-a(w(\theta))} \frac{\partial}{\partial \theta_2} e^{a(w(\theta))} = \mathbb{E}_\theta \big( \sum_{i=1}^{k} \frac{\partial w_i}{\partial \theta_2} t_i \big). \tag{2}$$

# 4 Random Samples

#### 4.0.1 The Delta Method

**Theorem 4** (**Delta Method, Theorem 4.14 in 541A notes, 5.5.24 in Casella and Berger**). Let $\theta \in \mathbb{R}$. Let $Y_1, Y_2, \ldots$ be random variables such that $\sqrt{n}(Y_n - \theta)$ converges in distribution to a mean zero Gaussian random variable with variance $\sigma^2 > 0$. Let $f : \mathbb{R} \to \mathbb{R}$. Assume that $f'$ exists and is continuous, and $f'(\theta) \neq 0$. Then

$$\sqrt{n}(f(Y_n) - f(\theta)) \xrightarrow{d} \mathcal{N}(0, \sigma^2 (f'(\theta))^2).$$

**Theorem 5** (**Second Order Delta Method, Theorem 4.17 in Math 541A Notes.**). Let $\theta \in \mathbb{R}$. Let $Y_1, Y_2, \ldots$ be random variables such that $\sqrt{n}(Y_n - \theta)$ converges in distribution to a mean zero Gaussian random variable with variance $\sigma^2 > 0$. Let $f : \mathbb{R} \to \mathbb{R}$. Assume that $f''$ exists and is continuous, $f'(\theta) = 0$ and $f''(\theta) \neq 0$. Then

$$n(f(Y_n) - f(\theta)) \xrightarrow{d} \sigma^2 \frac{1}{2} |f''(\theta)| \cdot \chi_1^2.$$