Statistical Theory Cheatsheet

Compiled by Gejun Zhu (zhug3@miamioh.edu) in preparation for the theory comprehensive exam by using William Chen's formula sheet template.

Last Updated August 25, 2015

Probability and Distributions

Section 1 - probability and distributions
Let $Y = g(X)$, where $g(x)$ is a one-to-one differentiable function.
$f_Y(y) = f_X(g^{-1}(y))|\frac{dx}{dy}|$ for $y \in S_Y$.
If X and Y are independent random variables, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.
If X and Y are independent random variables, then $\rho(X,Y) = 0$.
If X is a discrete random variable, then mgf of X is $M_X(t) = \sum P(X=k)e^{kt}$.
If $X \sim Gamma(a_1, b)$ and $Y \sim Gamma(a_2, b)$, $X + Y \sim Gamma(a_1 + a_2, b)$ if X and Y are independent.
If $X \sim Gamma(a_1, b)$, then $cX \sim Gamma(a_1, cb)$.
$X \sim Gamma(a, b)$, if $a = 1$ then $X \sim exponential(b)$; if $b = 2$ then $X \sim \chi^2_{(2a)}$.
If $X_i \sim Laplace(\mu, b)$ then $\frac{2\sum_{i=1}^n |X_i - \mu|}{b} \sim \chi^2(2n)$.
If $X \sim Laplace(0, b)$ then $|X|\ Exponential(b^{-1})$.
If $X \sim Exp(\lambda)$ then $X \sim Gamma(1, \lambda)$.
If $X \sim Gamma(k, \theta)$ where k is the shape parameter and θ is a scale parameter, then $E(X) = k\theta$, $var(X) = k\theta^2$; if $X \sim Gamma(\alpha, \beta)$ where α is the shape parameter. β is the rate parameter, then $E(X) = \frac{\alpha}{\beta}$, $var(X) = \frac{\alpha}{\beta^2}$.

Section 2 - Multivariate Distributions
Covariance: $EX = \mu_1$, $EY = \mu_2$, $Cov(X,Y) = E(XY) - \mu_1\mu_2$,
Correlation coefficient: $\rho = \frac{cov(X,Y)}{\sigma_1\sigma_2}$
$E(X_2) = E[E(X_2|X_1)]$,
$Var(X_2) = E[Var(X_2|X_1)] + Var(E(X_2|X_1)) \geq Var(E(X_2|X_1))$.
Moment generate function: $M = E(e^{tX})$, $\mu = E(X) = M'(0)$,
$\sigma^2 = E(X^2) - (EX)^2 = M''(0) - [M'(0)]^2$
$Z_n \sim \chi^2(n)$, $M_{Z_n} = (1-2t)^{-\frac{n}{2}}, t < \frac{1}{2}$, $w_n = \frac{Z_n}{n^2}$,
$M_{w_n}(t) = E[e^{tw_n}] = E[e^{t\frac{Z_n}{n^2}}] = M_{Z_n}(\frac{t}{n^2}) = (1 - \frac{2t}{n^2})^{-\frac{n}{2}}$ for $\frac{t}{n^2} < \frac{1}{2}$
Negative binomial distribution: $y = \#$ of failures before the r^{th} success.
$p(y) = \binom{y+r-1}{r-1}p^r(1-p)^y$
Poisson distribution: $y = \#$ of successes in a fixed length of time, $p(y) = \frac{\lambda^x e^{-\lambda}}{x!}$;
Gamma distribution: $y =$ waiting time required until the α^{th} success.
$f(y) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{\frac{-x}{\beta}}$. Special cases: (1) exponential distribution - $\alpha = 1, \beta = \frac{1}{\lambda}$; (2) Chi-square distribution - $\alpha = \frac{r}{2}$, $\beta = 2$.

Statistical Inference

Inequalities - Important Inequalities
Markov's Inequality: $u(X)$ non-negative, $E(u(X))$ exists, $P[u(X) \geq c] \leq \frac{E(u(X))}{c}$.
Chebyshev's Inequality: $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

Distributions : **Some facts**
Sample mean: $\bar{X} = \frac{\sum X_i}{n}$, Sample variance: $S^2 = \frac{\sum X_i^2 - n\bar{X}^2}{n-1}$, $E(S^2) = \sigma^2$; another unbiased estimator for σ^2 in normal distribution is $(\bar{x}^2 - \frac{S^2}{n})$ [also think about Y_n]
$E(\bar{X}) = \mu$, $var(\bar{X}) = \frac{\sigma^2}{n}$, $\bar{X} \sim N(\mu, \sigma^2/n)$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$, $\bar{X}$ & S^2 are indep.;
$\sum a_i X_i$ and $\sum b_i X_i$ are indep. iff $\sum a_i b_i = 0$;
$\frac{\sum(X_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{(\bar{X}-\mu)^2}{\sigma^2/n}$
$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ by CLT, $\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim T_{n-1}$

Order Statistics - **Order Statistics**
$g_k(y_k) = \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1 - F(y_k)]^{n-k}f(y_k)$, $a < y_k < b$; 0 elsewhere.

Section 4 - **Confidence Interval Estimation**
Error types: Type I - reject H_0 while H_0 is true, $P_\theta(X \in R)$; Type II - fail to reject H_0 while H_0 is false. $\alpha = P(Type\ I\ error)$, $\beta = P(Type\ II\ error)$, $Power = 1 - \beta$.
Level of significance: the probability of making type I error - reject H_0 when H_0 is true.
Power function: $K(\theta) = P_\theta[(X_1, X_2, ..., X_n) \in C|\theta]$, $\theta \in \omega_1$. The *power function* of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(X \in R)$. In other words, the power of a hypothesis test is the probability of rejecting H_0 when H_a is true. $Power = P(X \in R|\theta = \theta_a) = 1 - \beta$.
Level of significance : $\alpha = max_{\theta \in \omega_0} K(\theta)$ or
$\alpha = P(Type\ I\ error) = P(rejecting\ H_0\ when\ H_0\ is\ true) = P(X \in R|\theta = \theta_0)$
CI for difference in Means: $X_i \stackrel{iid}{\sim} N(\mu_1, \sigma^2)$, $Y_i \stackrel{iid}{\sim} N(\mu_2, \sigma^2)$, where σ^2 is unknown. $S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$, $\frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}} \sim T(n+m-2)$.
Note: $x > \theta \Rightarrow I(x - \theta)$

Section 5 - **Consistency and Limiting Distributions**
Convergence in Probability: $X_n \xrightarrow{P} X$ if for every $\epsilon > 0$,
$lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$ or ($lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$).
Degenerate r.v.: $p(x) = 1, if\ x = a; p(x) = 0,\ if\ x \neq a$ and
$F(x) = 0\ if\ x < a; F(x) = 1\ if\ x \geq a$. We write $X_b \xrightarrow{P} a$.
Consistency: The statistic T_n is a consistent estimator for θ iff $T_n \xrightarrow{P} \theta$. [find limit of the estimator $E(\hat{\theta}) \to \theta$ and $var(\hat{\theta}) \to 0$]
Convergence in Distribution: $X_n \xrightarrow{D} X$ if $lim_{n\to\infty} F_n(x) = F(x)$. F(x) is said to be the limiting distribution or asymptotic distribution of X;
Limiting distribution: $X_n \to X$ if and only if $lim_{n\to\infty} F_n(x) = F(x)$ where $F_n(x)$ is the cdf of X_n. $F(x)$ is said to be the limiting distribution of X_n.
Theorem 5.2.10: Suppose X_n has m.g.f. $M_{X_n}(t)$ that exists for $-h \leq x \leq h$ for all n. Let X has m.g.f. $M(t)$ which exists for $|t| \leq h_1 \leq h$. If $lim_{n\to\infty} M_{X_n}(t) = M(t)$ for $|t| \leq h_1$, then $X_n \xrightarrow{D} X$.
m.g.f technique: (1). $lim_{n\to\infty}(1 + \frac{b}{n} + \frac{\phi(n)}{n})^{cn} = lim_{n\to\infty}(1 + \frac{b}{n})^{cn} = e^{bc}$ where b and c are constants and $lim_{n\to\infty}\phi(n) = 0$. (2). $e^x = 1 + x + \frac{x^2}{2} + ... + \frac{x^m}{m!} + ...$;
CLT: $X_1, X_2, ..., X_n \stackrel{iid}{\sim} f(x)$ with mean μ and variance σ^2, $y_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} z$, where $z \sim N(0,1)$.

Section 6 - **Maximum Likelihood Estimation**
Consistent: If there is a unique solution to the likelihood equation $\frac{\partial}{\partial\theta}L(\theta) = 0$, then $\hat{\theta} \xrightarrow{P} \theta$ ($\hat{\theta}$ is consistent for θ).
Score function: $\frac{\partial ln(f(x;\theta))}{\partial\theta}$;
Fisher Information: $I(\theta) = var(\frac{\partial ln(f(x;\theta))}{\partial\theta}) = -E(\frac{\partial^2 ln(f(x;\theta))}{\partial\theta^2})$;
Efficient: y is unbiased for θ, y is efficient for θ iff $var(y) = [nI(\theta)]^{-1}$. In general, $var(y) \geq [nI(\theta)]^{-1}$ where y is unbiased for θ.
Efficiency: The efficiency of an unbiased estimator is given by the ratio $\frac{RCLB}{var(\hat{\theta})}$, where $RCLB = [nI(\theta)]^{-1}$.
Relative Efficiency: Relative efficiency of unbiased estimators $\hat{\theta}_1$ to $\hat{\theta}_2$ is $\frac{var(\hat{\theta}_2)}{var(\hat{\theta}_1)}$.
If θ_2 is biased, then relative efficiency is $\frac{var(\theta_2) + [bias(\theta_2)]^2}{var(\theta_1)}$ where $bias(\theta_2) = E(\theta_2) - \theta$.
Theorem 6.1.2: Suppose $\hat{\theta}$ is the MLE of θ and $g(\theta)$ is a function of θ. Then MLE of $g(\theta)$ is $\hat{g(\theta)} = g(\hat{\theta})$.
Theorem 6.2.1 (Rao-Cramer Lower Bound): $X_1, X-2, ..., X_n \stackrel{iid}{\sim} f(x;\theta)$ for $\theta \in \Omega$. Let $y = u(X_1, X_2, ..., X_n)$ be a statistic with mean $E(Y) = k(\theta)$. Then $var(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}$.
MVUE: $\hat{\theta} = u(X_1, X_2, ..., X_n)$ is a minimum variance unbiased estimator for θ iff $E(\hat{\theta})$ and $var(\hat{\theta})$ is less than or equal to the variance of every other unbiased estimator.
Theorem: If $\hat{\theta}$ is asymptotically unbiased for θ and $var(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta} \xrightarrow{P} \theta$.

Section 7 - **Measure of Quality of Estimators**
MVUE: $\hat{\theta} = u(X_1, X_2, ..., X_n)$ is a minimum variance unbiased estimator for θ iff $E(\hat{\theta}) = \theta$ and $var(\hat{\theta})$ is less than or equal to the variance of every other unbiased estimator. Relative efficiency of MVUE to any other unbiased estimator must be ≥ 1. MUVE would be consistent if $var(\hat{\theta}) \to 0$ as $n \to \infty$.
Sufficient Statistics: Let $X_1, X_2, ..., X_n$ be a random sample from $f(x;\theta)$, $y_1 = u(X_1, X_2, ..., X_n)$ is sufficient for θ iff $\frac{\Pi(f(x_i;\theta))}{g_1(y_1;\theta)} = H(x_1, x_2, ..., x_n)$ where g_1 is a marginal pdf for y_1.
Factorization Theorem: $y_1 = u(x_1, x_2, ..., x_n)$ is sufficient for θ iff $\Pi_{i=1}^n f(x_i;\theta) = k_1(u(x_1, x_2, ..., x_n);\theta)k_2(x_1, x_2, ..., x_n)$ where $k_2(x_1, x_2, ..., x_n)$ does not depend on θ.
Theorem 7.3.2: $X_1, X_2, ..., X_n \stackrel{iid}{\sim} f(x;\theta)$. If a sufficient statistic $y_1 = u(X_1, X_2, ..., X_n)$ for θ exists and if an MLE $\hat{\theta}$ exists uniquely, then $\hat{\theta}$ is a function of y_1.
Theorem: Suppose $y_1 = u(X_1, X_2, ..., X_n)$ is a sufficient statistic for θ. Let $z = u(y_1)$ be a 1-to-1 transformation not involving θ. z is also sufficient for θ.
Rao-Blackwell Theorem: (1). $y_1 = u_1(x_1, x_2, ..., x_n)$ be sufficient for θ; (2) $y_2 = u_2(x_1, x_2, ..., x_n)$ be unbiased for θ; (3). $\phi(y_1) = E(y_2|y_1)$. Then, (1) $\phi(y_1)$ is a statistic; (2) $\phi(y_1)$ is a function of y_1 alone; (3) $\phi(y_1)$ is unbiased for θ; (4). $\phi(y_1)$ has variance $< \sigma^2_{y_2}$.
Completeness: Suppose $Z \sim h(z;\theta)$, a member of a family of p.d.f's (p.m.f's): $\{h(z;\theta), \theta \in \Omega\}$. If $E(u(z)) = 0, \forall\theta \in \Omega$ implies that $u(z) = 0$ except on a set of points that has probability 0 for each $h(z;\theta), \theta \in \Omega$, then the family $\{h(z;\theta), \theta \in \Omega\}$ is called a complete family of density (mass) functions.
Lehmann and Scheffe (MVUE): Let $Y_1 = u_1(X_1, X_2, ..., X_n)$ be sufficient for θ and $\{g_1(y_1, \theta) : \theta \in \Omega\}$ be a complete family of densities (or p.m.f's). If there is a function if Y_1 which is unbiased for θ, then this function of Y_1 is the unique MVUE for θ.
Find MVUE: (1) Find a sufficient statistic t; (2) show that the family of distributions of t is complete [shortcut - theorem 7.5.2]; (3) Find a crude unbiased estimator; (4) evaluate
Exponential Family: $f(x;\theta) = exp\{p(\theta)k(x) + \delta(x) + q(x)\}$ where $\delta(x)$ dose not depend on θ, $p(\theta)$ is nontrivial continuous, $k'(x) \neq 0$.
$X_1, X_2, ..., X_n \stackrel{iid}{\sim} f(x;\theta)$, a regular case of exponential class with $y_1 = \sum k(x_i)$.
Then, (1) $g_1(y_1;\theta) = R(y_1)exp\{p(\theta)y_1 + nq(\theta)\}$; (2) $E(y_1) = -n\frac{q'(\theta)}{p'(\theta)}$; (3) $var(y_1) = n\frac{1}{[p'(\theta)]^3}\{p''(\theta)q'(\theta) - q''(\theta)p'(\theta)\}$
Theorem 7.5.2: $X_1, X_2, ..., X_n \stackrel{iid}{\sim} f(x;\theta)$, a regular case of exponential class with $\Omega = \{\theta : \gamma < \theta < \delta\}$, $y_1 = \sum k(x_i)$ is **sufficient** for θ and the family $\{g_1(y_1;\theta) : \gamma < \theta < \delta\}$ is **complete**.
Theorem 7.4.1: $X_1, X_2, ..., X_n$, a random sample from $f(x;\theta)$, $y_1 = u_1(X_1, X_2, ..., X_n)$ complete sufficient for θ and $E(\phi(y_1)) = \alpha(\theta)$, then $\phi(y_1)$ is unique MVUE for $\alpha(\theta)$.
unique MVUE = sufficient + complete (Theorem 7.5.2) + unbiased

Section 8 **Most powerful tests**
Neyman-Pearson Theorem: A **best critical region** of size α for testing $H_0 : \theta = \theta'$ v.s $H_1 : \theta = \theta''$ (both simple) is such that
(1) $\frac{L(\theta'; x_1, x_2, ..., x_n)}{L(\theta''; x_1, x_2, ..., x_n)} \leq k$ for each $(x_1, x_2, ..., x_n) \in C$
(2) $\frac{L(\theta'; x_1, x_2, ..., x_n)}{L(\theta''; x_1, x_2, ..., x_n)} \geq k$ for each $(x_1, x_2, ..., x_n) \in C^c$
(3) $P((x_1, x_2, ..., x_n) \in C; H_0) = \alpha$
Uniformly Most Powerful Test: C is a uniformly most powerful critical region of size α if C is a best critical region of size α for testing H_0 against **each** simple hypothesis in H_1. Because the critical region C defines a test that is most powerful against each simple alternative H_1, this is a uniformly most powerful test, and C is a uniformly most powerful critical region of size α.
Likelihood Ratio Test: Let (1) $L(\hat{\omega})$ denote the maximum of the likelihood function with respect to θ when θ is in the null parameter space ω. (2) $L(\hat{\Omega})$ denote the maximum of the likelihood function with respect to θ when θ is in the entire parameter space Ω. Then, the likelihood ratio is the quotient:

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{L(\hat{\theta}_0|x)}{L(\hat{\theta}|x)}.$$

p.s: MLE estimated μ and σ^2 for normal distribution: $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$; $\hat{\mu} = (n+m)^{-1}\{\sum x_i + \sum y_i\}$, $\hat{\sigma}^2 = (n+m)^{-1}\{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2\}$.
Neyman-Pearson Lamma: If a critical value is chosen so that $P_{\theta_0}(\Lambda \leq c) = \alpha$, then the test with decision rule

$$\text{Reject } \theta = \theta_0 \text{ in favor of } \theta = \theta_1 \text{ when } \Lambda \leq c$$

is a uniformly most powerful test of size α.