Statistics with Recitation
National Taiwan University
Fall 2020
Shiu-Sheng Chen

Midterm Brief Review

Last Edited: December 05, 2020

Hsien-Chen Chu T09303304 Econ1

Literature: S.S.Chen: Probability and Statistical Inference with R, **ch6-10**, 2019.
I'd like to reiterate that this is merely a BRIEF review of midterm contents, so it leaves most of the proofs and inferences for you to revisit the textbook if needed.
Important formulas and features are listed sequentially.

1. **Normal Distribution:** $\sim N$

   (a) **Normal:**

   $$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \Rightarrow X \sim N(\mu, \sigma^2)$$

   (b) **Standard Normal:** def $z = \frac{x-\mu}{\sigma}$, then:

   $$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \Rightarrow Z \sim N(0, 1)$$

   (c) Linear Trans.: given $X \sim N(\mu, \sigma^2)$ and $aX+b$, then: $aX+b \sim N(a\mu+b,\ a^2\sigma^2)$

   (d) i. i.i.d sum: $Y = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

   ii. i.i.d mean: $W = \frac{Y}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N(\mu, \frac{\sigma^2}{n})$

   (e) General: re-def $W = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n \Rightarrow W \sim N(\sum_{i=1}^{n} \alpha_i \mu_i,\ \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2)$

2. **Chi-square Distribution:** $\sim \chi^2$

   (a) **Chi-square:**

   $$f(x) = \frac{x^{\frac{k}{2}-1}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} e^{-\frac{1}{2}x} \Rightarrow X \sim \chi^2(k)$$

   Degrees of freedom: **k**

   (b) MGF:

   $$X_i \sim \chi^2(k) \Rightarrow M_X(t) = (\frac{1}{1-2t})^{\frac{k}{2}}$$

   i. $E(X) = k$
   ii. $E(X^2) = k(k+2)$
   iii. $Var(X) = 2k$

(c) Chi-square sum: Suppose $X_i \sim \chi^2(k_i)$, then:

$$Y = \sum_{i=1}^{n} X_i \sim \chi^2(\sum_{i=1}^{n} k_i)$$

(d) **Standard Normal & Chi-square:** given $Z \sim N(0,1)$, then:

$$Z^2 \sim \chi^2(1)$$

Lemma
Given $\{Z_1, Z_2, \ldots, Z_n\} \sim^{i.i.d.} N(0,1),\ let\ X = \sum_{i=1}^{k} Z_i^2,\ then: X \sim \chi^2(k).$

3. **Student's t Distribution:** $\sim t$

(a) **Standard Normal, Chi-square & t:** given $Z \sim N(0,1),\ W \sim \chi^2(k)$, then:

$$X = \frac{Z}{\sqrt{\frac{W}{k}}} \sim t(k)$$

Degrees of freedom: **k**

(b) Features: (parameters)

$$t(k) = \frac{N(0,1)}{\sqrt{\frac{\chi^2(k)}{k}}}$$

i. $E(X) = E(Z\frac{1}{\sqrt{\frac{W}{k}}}) = E(Z)E(\frac{1}{\sqrt{\frac{W}{k}}}) = 0 \times E(\frac{1}{\sqrt{\frac{W}{k}}}) = 0$

ii. $Var(X) = \frac{k}{k-2}$, for $k > 2$

4. **F Distribution:** $\sim F$

(a) **F:**

$$X \sim F(k_1, k_2)$$

Degrees of freedom: k_1, k_2

(b) **Chi-square & F:** given $X_1 \sim \chi^2(k_1),\ X_2 \sim \chi^2(k_2)$, then:

$$X = \frac{X_1/k_1}{X_2/k_2} \sim F(k_1, k_2)$$

Remarks: F distribution is the ratio of 2 Chi-square distributions, each divided by its own degrees of freedom. And, suppose $Y = \frac{1}{X}$, then: $Y \sim F(k_2, k_1)$ by definition.

(c) **t & F:** given $X \sim t(k)$, then:

$$X^2 \sim F(1, k)$$

5. **Random Samples & Descriptive Stats:**

   (a) Random Samples: $\{X_i\}_{i=1}^n \sim^{i.i.d.} (\mu, \sigma^2)$

   (b) Statistics:

      i. Sample mean: $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$

      ii. Sample variance: $S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}$

   (c) Suppose $\{X_i\}_{i=1}^n \sim^{i.i.d.} N(\mu, \sigma^2)$, then:

$$\begin{cases} \bar{X}_n \sim N(\mu, \frac{\sigma^2}{n}) \\ \frac{\sum_i (X_i - \bar{X}_n)^2}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1) \\ \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t(n-1) \end{cases} \tag{1}$$

   Recall: 1.(d) i. $\rightarrow$ i.i.d. mean, and $\bar{X}_n \perp S_n^2$ (indep.).

   (d) Suppose $\{X_i\}_{i=1}^m \sim^{i.i.d.} N(\mu_X, \sigma_X^2)$ and $\{Y_i\}_{i=1}^n \sim^{i.i.d.} N(\mu_Y, \sigma_Y^2)$, then:

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1,\ n-1)$$

   Degrees of freedom: $k_X = m-1,\ k_Y = n-1$

6. **Convergence:**

   (a) Markov Inequality: given a random variable X, $\forall m > 0$,

$$P(X \geq m) \leq \frac{E(X)}{m}$$

   (b) **Chebyshev Inequality:** given a random variable $Y \sim (E(Y), Var(Y))$, $\forall \varepsilon > 0$,

$$P(|Y - E(Y)| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}$$

   (c) **Convergence:**

      i. Converge in Probability:

$$\begin{cases} \lim_{n\to\infty} P(|Y_n - c| < \varepsilon) = 1 \Rightarrow Y_n \xrightarrow{p} c \\ \lim_{n\to\infty} P(|Y_n - Y| < \varepsilon) = 1 \Rightarrow Y_n \xrightarrow{p} Y \end{cases} \tag{2}$$

      ii. Converge in Distribution:

$$\begin{cases} \lim_{n\to\infty} F_n(y) = F_Y(y) \Rightarrow Y_n \xrightarrow{d} Y \\ \lim_{n\to\infty} M_{Y_n}(t) = M_Y(t) \Rightarrow Y_n \xrightarrow{d} Y \end{cases} \tag{3}$$

iii. Converge in Mean Square:

$$\begin{cases} E[(Y_n - c)^2] \to 0 \ as \ n \to \infty \Rightarrow Y_n \xrightarrow{ms} c \\ E[(Y_n - Y)^2] \to 0 \ as \ n \to \infty \Rightarrow Y_n \xrightarrow{ms} Y \end{cases} \tag{4}$$

(d) Relationship:

$$\lim_{n\to\infty} E(Y_n) = c, \ \lim_{n\to\infty} Var(Y_n) = 0$$
$$\Leftrightarrow Y_n \xrightarrow{ms} c \Rightarrow Y_n \xrightarrow{p} c$$

Remarks: The first half of cond. is an "iff", and the second half is an "if".

(e) **WLLN** (Weak Law of Large Numbers)**:** given $\{X_i\}_{i=1}^n$, $Var(X_i) < \infty$. Let $\bar{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$, then:

$$\bar{X_n} \xrightarrow{p} E(X_1)$$

Remarks: With a large scale of samples($n \to \infty$), the sample mean($\bar{X_n}$) will be close to the expected value($= E(X_1) = \mu$).

i. k-th moments:
given $E(X_1^r) < \infty$, then: $\frac{\sum_i X_i^r}{n} \xrightarrow{p} E(X_1^r)$
given $E(X_1Y_1) < \infty$, then: $\frac{\sum_i X_iY_i}{n} \xrightarrow{p} E(X_1Y_1)$

ii. Application: Suppose $W_n \sim Binomial(n, \mu)$, let $Y_n = \frac{W_n}{n}$, then: $Y_n \xrightarrow{p} \mu$

(f) **CLT** (Central Limit Theorem) **:** given $\{X_i\}_{i=1}^n$, $Var(X_i) < \infty$, $E(X_1) = \mu < \infty$, $Var(X_1) = \sigma^2 < \infty$, then:

$$\frac{\bar{X_n} - E(\bar{X_n})}{\sqrt{Var(\bar{X_n})}} = \frac{\bar{X_n} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \xrightarrow{d} N(0,1)$$

(g) Other Convergences:

i. **CMT** (Continuous Mapping Theorem): $g(\cdot)$ *conti.*, then: $g(X_n) \xrightarrow{p} g(X)$

ii. $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$, $X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{p} c$

iii. **Slutsky's Theorem:** given $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{p} c$, then:

$$[X_n + Y_n \ / \ X_nY_n \ / \ \frac{X_n}{Y_n}] \xrightarrow{d} [X + c \ / \ cX \ / \ \frac{X}{c}, \ c \neq 0]$$

iv. The Delta Method

7. **Point Estimation:**

(a) **MME** (Method of Moments Estimators)**:** given pdf $f(x, \theta_1, \theta_2, \ldots, \theta_k)$, solve:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^j = m_j(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \xrightarrow{p} m_j(\theta_1, \theta_2, \ldots, \theta_k), \ j = 1, 2, \ldots, k$$

Remarks: Sample's j-th moments = Population's j-th moments(by WLLN).

(b) **MLE** (Maximum Likelihood Estimator)**:**
Likelihood function

$$\mathcal{L}(\theta) = \prod_i f(x_i, \theta),\ \theta = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$$

→ Find MLE by solving FOC:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = 0 \ or \ \frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta} = 0$$

i. **Unbiased:**
Unbiased Estimator: $\hat{\theta}$, we expect that:

$$E(\hat{\theta}) = \theta$$

If not, then there exists **bias:** $B(\theta) = E(\hat{\theta}) - \theta$

Suppose $\{X_i\}_{i=1}^n \sim^{i.i.d.} N(\mu, \sigma^2)$, then:

$$\begin{cases} \bar{X} = \frac{\sum_{i=1}^n X_i}{n} \Rightarrow E(\bar{X}) = \mu \ (unbiased) \\ S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \Rightarrow E(S^2) = \sigma^2 \ (unbiased) \\ \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \Rightarrow E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \ (biased) \end{cases} \tag{5}$$

Remarks:
The difference between S^2 and $\hat{\sigma}^2$ is the denominator($n-1$ vs n). Apparently, $\bar{X}$ and S^2 are unbiased estimators of μ and σ^2, respectively, while $\hat{\sigma}^2$ is a biased one of σ^2, with its $B(\hat{\sigma}^2) = -\frac{1}{n}\sigma^2$.

→ **MVUE** (Minimum Variance Unbiased Estimator): $\hat{\theta}$ is MVUE of θ
$\Leftrightarrow E(\hat{\theta}) = \theta,\ \forall \theta$
$\Leftrightarrow Var(\hat{\theta}) \leq Var(\hat{\theta}^*),\ \forall \hat{\theta}^*,\ E(\hat{\theta}^*) = \theta$

ii. **Efficient:**
MSE (Mean Squared Error)**:**

$$MSE(\hat{\theta}) \equiv E[(\hat{\theta} - \theta)^2] \Rightarrow MSE(\hat{\theta_n}) = Var(\hat{\theta_n}) + (B(\theta))^2$$

Remarks: An estimator that has a smaller MSE is the more efficient one, whether it is unbiased. Furthermore, from (i)'s example, we know $MSE(S^2) = \frac{2\sigma^4}{n-1} + 0 = \frac{2\sigma^4}{n-1}$ and $MSE(\hat{\sigma}^2) = \frac{2(n-1)\sigma^4}{n^2} + (-\frac{\sigma^2}{n})^2 = \frac{(2n-1)\sigma^4}{n^2}$

$$\Rightarrow MSE(\hat{\sigma}^2) < MSE(S^2)$$

Thus, the biased $\hat{\sigma}^2$ is a more efficient estimator than the unbiased S^2.

iii. **Consistent:** $\hat{\theta}_n$ is a consistent estimator of θ when:

$$\hat{\theta}_n \xrightarrow{p} \theta$$

Remarks: A subscript n reminds us that this feature is related to the sample size(similar to WLLN). Again, from (i)'s example, we eventually know $\bar{X}_n$ is a consistent estimator of μ, and both S_n^2 and $\hat{\sigma}_n^2$ are consistent estimators of σ^2(proved by WLLN and CMT).

MSE Consistent: $\hat{\theta}_n$ is a MSE consistent estimator of θ when:

$$\hat{\theta}_n \xrightarrow{ms} \theta$$

Also, consider

$$\hat{\theta}_n \xrightarrow{ms} \theta \Rightarrow \hat{\theta}_n \xrightarrow{p} \theta$$

Remarks: MSE consistent is based on the idea of mean square convergence. From 6.(d) *Relationship* $\rightarrow$ If $\hat{\theta}_n$ is MSE consistent, then it is a consistent estimator. Thus, to evaluate whether $\hat{\theta}_n$ is a consistent estimator, we can simply check if $\{\lim_{n\to\infty} E(\hat{\theta}_n) = \theta, \ \lim_{n\to\infty} Var(\hat{\theta}_n) = 0\}$ were satisfied.

Asymptotically Unbiased: $\hat{\theta}_n$ is asymptotically unbiased when:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

And,

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \lim_{n\to\infty} \theta = \theta$$

Remarks: If $\hat{\theta}_n$ is unbiased, then it is also asymptotically unbiased.