Econometrics Year 1 Cheat-sheet
Julian Budde, Page 1 of 2

1 Probability Theory

Thm 1.3 (LTP): $A_1,...,A_2$ partition of S and $B \subset S$, then $P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$.

Thm 1.4(Bayes' rule): $A_1, A_2, \ldots$ partition of S, B any set. Then for each $i = 1, 2, \ldots$, $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}$.

Def. 1.5 (Sigma Algebra): Collection of subsets of S is a *sigma algebra* $\mathcal{B}$ if it satisfies: (1) $\emptyset \in \mathcal{B}$, (2) If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$, and (3) if $A_1, A_2, \ldots \in \mathcal{B}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$.

Mutual independence $\Rightarrow$ *pairwise independence*, but not $\Leftarrow$.

2 Random Variables

Thm 2.5 (Jensen's inequalities): Suppose $g(x)$ convex, then $E(g(X)) \geq g(E(X))$ if existent. Strict unless X degenerate or g linear.

Def 2.10 (MGF): $X \sim F_X$, $t \in \mathbb{R}$. Then $M_X(t) = E(e^{tX})$ given it exists in some neighborhood of 0.

Thm 2.7: If $M_X(t)$ exists, then $E(X^n) = \frac{\partial^n}{\partial t^n} M_X(0)$.

3 Multivariate Distributions

Def 3.1: n-dimensional rvec is $f : S \to \mathbb{R}^n$.

3.1 Bivariate Random Vectors

Define probability functions on Borel sigma algebra of $\mathbb{R}^2$.

Need to assume $E(|g(X,Y)|) < \infty$.

Joint $\Rightarrow$ Marginal: $F_X(x) = lim_{y\to\infty} F_{X,Y}(x,y)$ and $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,v)dv$.

3.2 Continuous Distributions

Conditional Expectation: $E(g(Y)|X = x) = \sum_{y\in(Y)} g(y) f_{Y|X}(y|x)$ or $= \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x)dy$.

Thm 3.1 (LIE): Y, X rvs, then $E(Y) = E_X(E_{Y|X}(Y|X))$.

Law of iterated variance: $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$.

3.3 Independence

Def 3.4: (X,Y) rvec, X, Y *independent* if $\forall x \in \mathbb{R}, y \in \mathbb{R}$ we have $f_{X,Y}(x,y) = f_X(x) f_Y(y)$.

Thm 3.2: X, Y independent $\Leftrightarrow$ for any two bounded $g, h : \mathbb{R} \to \mathbb{R}$ we have $E(g(X)g(Y)) = E(g(X))E(h(Y))$.

Thm 3.3: X, Y independent, $g(X)$ and $g(Y)$ independent.

4 Sampling

4.1 Distribution of the t-ratio

With $\{X_i\}_{i=1}^{\infty}$ rs of $X_i \sim N(\mu, \sigma^2)$ we have $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma/} \sim N(0,1)$. Then the *t-ratio* $\frac{\overline{X}_n - \mu}{\frac{1}{\sqrt{n}} S_n} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{S_n^2/\sigma^2}} \sim t_{n-1}$.

5 Asymptotic Theory

5.1 Inequalities

Thm 5.1 (Markov's Inequality): X r.v., $g : \mathbb{R} \to [0,\infty)$, then $\forall \epsilon > 0$, $P(g(X) > \epsilon) \leq \frac{E(g(X))}{\epsilon}$.

Cor 5.1 (Chebyshev's Inequality): X r.v., then $\forall \epsilon > 0$, $P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$.

5.2 Modes of Convergence

Def 5.2: $plim_{n\to\infty} X_n = X \leftrightarrow \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$.

Def 5.4: $\{X_n\}_{n=1}^{\infty}$ converges in *distribution* to X $\leftrightarrow \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for every continuity point of x of $F_X(\cdot)$.

Def 5.5: $\{X_n\}_{n=1}^{\infty}$ converges in *mean square* to $X \leftrightarrow \lim_{n\to\infty} E[(X_n - X)^2] = 0$.

Thm 5.2: $X_n \xrightarrow{m.s.} \Rightarrow X_n \xrightarrow{p} X$. Proof by Chebyshev's inequality. The reverse is not true, consider $X_n \in 0, \sqrt{n}$ with probabilities $1 - 1/n, 1/n$.

Thm 5.3: $X_n \xrightarrow{p} \Rightarrow X_n \xrightarrow{d} X$. Proof uses definition of $\xrightarrow{p}$ and continuity. The reverse is generally *not true*, consider $X_n = Z \sim N(0,1)$ and $X, Z \sim N(0,1)$, have $F_{X_n}(x) = F_X(x)$ but $P(|Z - X| \geq \epsilon) > 0$. Exception: $X_n \xrightarrow{d} c \in \mathbb{R} \Rightarrow X_n \xrightarrow{p} c$.

5.3 Law of Large Numbers

Thm 5.6 (LLN i.i.d): $\{X_i\}_{i=1}^{\infty}$ seq. of iid rvs from F_X with $\mu = E(X)$ exist and finite. Then $\overline{X}_n \xrightarrow{p} \mu$.

Convergence Criteria: Need a combination of three assumptions: (1) finite mean and/or variance (no LLN for Cauchy), (2) bounds on asymptotic variance (e.g. not growing too fast with i), (3) restricted dependence.

5.4 Central Limit Theorem

Thm 5.7 (Lindeberg-Levy CLT): $\{X_i\}_{i=1}^{\infty}$ seq. of iid rvs from F_X, μ and σ^2 finite. Then $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Thm 5.9 (Berry-Esseen): $\{X_i\}_{i=1}^{\infty}$ seq. of iid rvs from F_X, μ and σ^2 finite and $\lambda = E(|X - E(X)|^3)$ exist and finite. Let $Z \sim N(0,1)$. Then $\left|P\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq x\right) - P(Z \leq x)\right| \leq \frac{C\lambda}{\sigma^3\sqrt{n}}$.

5.5 Convergence of Random Vectors

Def 5.7: $\mathbf{X}_n \xrightarrow{p} \mathbf{X} \leftrightarrow \lim_{n\infty} P(\|\mathbf{X}_n - \mathbf{X}\| < \epsilon) = 1$.

Def 5.8: $\mathbf{X}_n \xrightarrow{ms} \mathbf{X} \leftrightarrow \lim_{n\infty} E(\|\mathbf{X}_n - \mathbf{X}\|^2) = 0$.

Def 5.9: $\mathbf{X}_n \xrightarrow{d} \mathbf{X} \leftrightarrow \lim_{n\infty} F_{\mathbf{X}_n}(x) = F_{\mathbf{X}}(x)$ for every continuity point x of $F_{\mathbf{X}}(\cdot)$.

Thm 5.10 (Cramér-Wold): $\{\mathbf{X}_n\}_{n=1}^{\infty}$ seq. of K-dimensional random vectors. Then, $\forall \lambda \in \mathbb{R}^K$ we have $\lambda'\mathbf{X}_n \xrightarrow{d} \lambda'\mathbf{X} \leftrightarrow \mathbf{X}_n \xrightarrow{d} \mathbf{X}$.

5.6 CMT and Slutzky's

Thm 5.11 (CMT): Let $\{\mathbf{X}_n\}_{n=1}^{\infty}$ be a sequence of K-dim. rvecs $\mathbf{X}$ K-dim rvec, and $g : \mathbb{R}^K \to \mathbb{R}$ with discontinuity points D such that $P(\mathbf{X} \in D) = 0$.

(a) $\mathbf{X}_n \xrightarrow{p} \mathbf{X} \Rightarrow g(\mathbf{X}_n) \xrightarrow{p} g(\mathbf{X})$.

(b) $\mathbf{X}_n \xrightarrow{d} \mathbf{X} \Rightarrow g(\mathbf{X}_n) \xrightarrow{d} g(\mathbf{X})$.

Implication: Sums and products of convergent sequences converge. Does *not* hold for *mean square* convergence.

Thm 5.12 (Slutzky's): X_n, Y_n seq of rvs with $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c \in \mathbb{R}$, then $X_n + Y_n \xrightarrow{d} X + c$ and $X_n Y_n \xrightarrow{d} cX$, and if $c \neq 0$, $X_n/Y_n \xrightarrow{d} X/c$.

Extension to rvecs: $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ and $\mathbf{Y}_n \xrightarrow{p} \mathbf{C} \in \mathbb{R}^{K\times K}$, $\mathbf{C}$ invertible, then $\mathbf{Y}_n^{-1}\mathbf{X}_n \xrightarrow{d} \mathbf{C}^{-1}\mathbf{X}$.

Example CMT: $\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}\right)^2 \xrightarrow{d} N(0,1)^2 = \chi_1^2$.

Thm 5.13 (Delta-Method): X_n seq of rvs with LL-CLT applying. $g : \mathbb{R} \to \mathbb{R}$ *continuously diff.* at μ with $g'(\mu) \neq 0$. Then $\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} N(0, g'(\mu)^2\sigma^2)$. Proof: CMT and Slutzky's applied to Taylor's/intermediate value theorem.

5.7 Interval Estimation

Suppose $\{X_i\}_{i=1}^{n}$ is a seq of iid random variables with μ, σ^2 finite. Then an asymptotically valid CI for μ is given by

$$CI = \left[\overline{X}_n \pm \frac{z_{1-\alpha/2}}{\sqrt{n}} S_n\right]$$

where S_n is a consistent estimator of σ and $P(\mu \in CI) \to 1 - \alpha$. Proof: CLT, CMT, Slutzky.

5.8 Moment-Based Estimation

Parameter of interest: $\theta = h(E(g(X)))$ (simple case: X, θ scalars and $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ cont. diff.).

Moment-based estimator: $\hat{\theta}_n = h\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right)$. *Consistency* follows from LLN and CMT.

Large-sample distribution: If $Var(g(X)) < \infty$ CLT applies so $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i) - E(g(X))\right) \xrightarrow{d} N(0, Var(g(X)))$. By the *delta-method* if $h'(g(E(X))) \neq 0$ we have

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d}= N(0, h'(E(g(X))^2 Var(g(X)))).$$

6 Maximum Likelihood Estimation

Def 6.1 (likelihood function): $L_n(\theta) = \prod_{i=1}^{n} f(\mathbf{x}_i; \theta)$.

Equivalently, we define the *log-likelihood function* as $\log(L_n(\theta))$.

Thm 6.1: Suppose $\mathbf{X}$ is a random vector with pdf or pmf $f(\mathbf{x}; \theta_0)$. Then $E(\log(f(\mathbf{x}; \theta))) \geq E(ln(f(\mathbf{X}; \theta))), \forall \theta in \Theta$.

Thm 6.2: For $\tau(\theta)$ and $\hat{\theta}_n$ MLE of θ, we have $\tau(\hat{\theta}_n)$ is MLE of $\tau(\theta)$.

6.1 Distribution of the MLE

MLE Limit Distribution: $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, A^{-1}BA^{-1})$ with $A = E_\theta[-\frac{\partial^2}{\partial\theta\partial\theta'} ln(f(\mathbf{X}_i; \theta)]$ and $B = E_\theta[\frac{\partial}{\partial\theta} ln(f(\mathbf{X}_i; \theta)) \frac{\partial}{\partial\theta'} ln(f(\mathbf{X}_i; \theta))]$. B *var-cov* matrix of the score (since score mean zero by FOC). A is *Fisher information*.

Thm 6.3: Under weak reg. cond. (diff; interch. integr./diff.) we have $A = B$.

6.2 CRLB

We could try to define best estimator in terms of MSE. However, MSE might depend on θ (e.g. $\overline{X}_n$ vs. 1, the latter dominates for $\theta = 1$).

Progress: Focus on *unbiased* estimators and thus variance.

Thm 6.4: $\{X_i\}$ rs from $f(\mathbf{x}; \theta)$, $\hat{\theta}_n$ estimator of θ. Then under some reg conds $Var_\theta[\hat{\theta}_n] = \frac{(\frac{\partial}{\partial\theta} E_\theta[\hat{\theta}_n])^2}{nE_\theta[(\frac{\partial}{\partial\theta}\log(f(\mathbf{X}; \theta)))^2]}$.

Relative efficiency: $E_\theta[(\hat{\theta}_{1,n} - \theta)^2] \leq E_\theta[(\hat{\theta}_{2,n} - \theta)^2]$ for all $\theta \in \Theta$ and strict for some.

Asymptotic efficiency: Asymptotic distribution often implies *asymptotically unbiased*, efficiency than means attaining CRLB asymptotically.

7 Hypothesis Testing

7.1 Basics

Def 7.1: A *hypothesis* is a statement about the population distribution.

Def 7.2: H_0 (null hypothesis) and H_1 (alternative hypothesis) are the complementary hypothesis. We write $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$ with Θ_k mutually exclusive and exhaustive.

Simple hypothesis: Θ_0 is singleton. *Composite* hypothesis: Θ_1 more than one value.

Def 7.3: A *hypothesis test* is a rule when to reject H_0 (in favor of H_1) given the data. (*Accepting* H_0 is weird, e.g. what about $\theta_0 + \epsilon$?.)

8 Size and Power

T-I error: Reject H_0 although in fact true.

T-II error: *Not* reject H_0 although in fact false.

Error rates: *Probabilities* of making these errors (errors are random because they depend on the sample). Usually **trade-off** between I and II.

Def 7.4 (Power function): $\beta(\theta) = P_\theta(rejectH_0)$.

T-I error rate: $\beta(\theta)$ for any $\theta \in \Theta_0$.

T-II error rate: $\beta(\theta)$ for any $\theta \in \Theta_1$.

Def 7.5/7.6: For $\alpha \in [0,1]$, a test is *level* α if $\sum_{\theta\in\Theta_0} \beta(\theta) \leq \alpha$ (*size*: equality).

8.1 Test statistics and critical values

Goal: Derive statistic T and reject iff $T > c_\alpha$ controlling $sup_{\theta\in\Theta_0} P_\theta(T > c_\alpha) \Rightarrow$ need $F_T(t)$.

Ex 7.2 (Two-sided T): $X \sim N(\mu, \sigma^2)$ so $\theta = (\mu, \sigma^2)$. Test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ use $T = \frac{\sqrt{n}|\overline{X}_n - \mu_0|}{S_n}$ $|t_{n-1}|$ and reject for $T > c_\alpha = t_{n-1,1-\alpha/2}$. By construction $sup_{\theta\in\Theta_0} P_\theta(T > c_\alpha) = \alpha$. Note this holds for all $\sigma^2 \in \Gamma$ thus a test of level and size α.

Ex (One-sided T): $T = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n}$ with $c_\alpha = t_{n-1,1-\alpha}$ or $Z = -T$ and c_α unchanged (symmetry). Intuition: want to reject for large $\mu > \mu_0$ (right-sided).

Deriving $\beta(\theta)$: (1) add and subtract (true) μ, (2) look at behavior as μ changes.

Def (p-value): For any realization T^*, $p^* = inf\{p \in [0,1] : T^* > c_p\}$. Intuition: smallest α for which we would still reject.

Under H_0, $p \sim Unif[0,1]$ (require $P(p^* < \alpha) = \alpha$, i.e. want $Pr_\theta(rejectH_0) < \alpha$), but holds $\forall \alpha$.

p-value with simple H_0: If F_0 is strictly increasing, $p* = 1 - F_0(T*)$ (again: $p_{H_0} Unif[0,1]$).

With parametric distributions with multiple parameters (e.g. $N(\mu, \sigma^2)$) usually fix one parameter (e.g. σ^2) resulting in simple test but technically *composite* H_0.

8.2 Hypothesis Testing and CIs

Test-inversion: Assume test $H_0 : \theta = \theta_0$ (note: this is some H_0) and have test s.t. $P_{\theta_0}(rejectH_0) = \alpha$ (size α). Assume can perform for any $\theta_0 \in \Theta$. Then we have $CS = \{\theta_0 \in \Theta : notrejectH_0 : \theta = \theta_0\}$ with $P_\theta(\theta \in CS) = 1 - \alpha$ (*true* θ).

We can also do the reverse: From any CS with coverage rate $1 - \alpha$ can construct size α test as *reject* $\Leftrightarrow \theta_0 \notin CS$.

Ex. one-sided CI: Testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ (or $H_0 : \mu \leq \mu_0$) for normal case we have $CS = \{\mu_0 \in [\overline{X}_n - \frac{t_{1-\alpha,n-1}}{\sqrt{n}} S_n, \infty)\}$.

8.3 Asymptotic Approximations

Asymptotic argument: No parametric model $f_X(x; \theta)$, but, e.g., moments: $H_0 : E(X) = \mu$. $T \xrightarrow{d} |N(0,1)|$ and we can use $\Phi^{-1}(x)$ to control α asymptotically. In particular, $P(T > z_{1-\alpha/2}) \to \alpha$ under H_0.

Hypotheses: Set of distributions $\mathcal{P}$ with $\mathcal{P}_0 \subset \mathcal{P}$ set of distributions consistent with H_0.

Def 7.7 (Asymptotic power function): $\beta^a(P) = lim_{n\to\infty} \beta_n(P)$.

Def 7.8/7.9: test with $\beta^a(P)$ is *asymptotic level* α if $sup_{P\in P_0} \beta^a(P) \leq \alpha$ (*size*: equality).

Def 7.10: Test *consistent* against alternative $P \in P_1$ if $\beta^a(P) = 1$.

Example: $\mathcal{P} = \{P : E(X), E(X^2) < \infty\}$ and $\mathcal{P}_0 = \{P : E(X) = 1\} \subset \mathcal{P}$ and $\mathcal{P}_1 : \{P : E(X) \neq 1\} \subset \mathcal{P}$.

Problem: $\beta^a(P)$ might not be informative about finite sample (e.g. $H_0 : \mu = \mu_0 + \epsilon$).

Distributions

Normal: $E(X) = \mu$, $Var(X) = \sigma^2$. Sum of two independent Normals is Normal.

MVN: $\sim N(\mu, \Sigma)$. Any linear combinations are Normal. $(X,Y) \sim N(\mu, \Sigma)$, then $X \perp\!\!\!\perp Y \Leftrightarrow Cov(X,Y) = 0$.

Uniform: $X \sim Unif(a,b)$, $F_X(x) = \frac{x-a}{b-a}$, $f_X(x) = \frac{1}{b-a}$, $E(X) = \frac{1}{2(b-a)}$, $Var(X) = \frac{1}{12}(b-a)^2$. $\hat{b}_{MLE} = max\{X_1, \ldots, X_n\}$ (min for a); $\hat{b}_{MM} = 2\overline{X}_n$.

Uniform Order Statistics: $U_{(k)} \sim Beta(k, n+1-k)$ with $E(U_{(k)}) = \frac{k}{n+1}$. **Exponential**: $X \sim Expo(\theta)$ then $E(X^k) = k!\theta^k$, so $E(X) = \theta$ and $Var(X) = \theta^2$. $\hat{\theta}_{MLE} = \overline{X}_n$.

Pareto: $X \sim Pareto(\alpha)$ then $E(X^k)$ only exists if $\alpha > k$. Given that, $E(X) = \frac{\alpha}{\alpha-1}$ and $Var(X) = \frac{\alpha}{(1-\alpha)^2(\alpha-2)}$. $Y = log(X)$ $Expo(\alpha)$. 20-80 rule: $\alpha = \frac{\ln 5}{\ln 4} \approx 1.16$.

t: If $Y \sim N(0,1)$ and $Z \sim \chi_{n-1}^2$ and $X \perp\!\!\!\perp Z$ then $\frac{N}{Z}$ t_{n-1}.

Cauchy: $X, Y \sim N(0,1)$ with $X \perp\!\!\!\perp Y$, then $\frac{X}{Y}$ $Cauchy(0,1)$. Expectation and variance undefined. $X \sim Cauchy(0,1)$ then $X \sim t_1$.