Midterm Review

450A Political Methodology I, handout (Lectures 1 to 4-3)

1 Probability (background)

- Probability rules; conditional probability
- Bayes' rule

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$$

- Independence; conditional independence

2 Random variables

- PDF, PMF, CDF
- Expectation (discrete and continuous)

$$\mathbb{E}[X] = \sum_x x\, P(X = x), \qquad \mathbb{E}[X] = \int x f(x)\, dx$$

- Variance

$$\sigma_X^2 = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- Normal distribution
- Conditional expectation; law of iterated expectations, $\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}[Y \mid X]\big]$
- Conditional variance
- Covariance

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- Correlation

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

3 Point estimation (Lecture 2-1)

- Estimands, estimators, and estimates
- Unbiasedness and bias, $\mathbb{E}[\hat{\theta}] - \theta$
- Efficiency

- Mean squared error

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \big(\mathbb{E}[\hat{\theta}] - \theta\big)^2 + \mathrm{Var}(\hat{\theta})$$

- Consistency; convergence in probability
- Weak law of large numbers (statement)
- Asymptotic normality; convergence in distribution
- Central limit theorem (statement)

4 Interval estimation (Lecture 2-2)

- Interpreting a confidence interval: coverage over repeated samples
- Estimating the population variance

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

- Standard error of $\bar{X}_n$

$$SE(\bar{X}_n) = S_n/\sqrt{n}$$

- t distribution and t statistic (exact if the population is normal)

$$T_n = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} = \frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \sim \tau_{n-1}$$

- Confidence intervals: exact with a normal population, asymptotic otherwise

$$\big[\bar{X}_n - t_{n-1,\,0.975}\,SE(\bar{X}_n),\ \bar{X}_n + t_{n-1,\,0.975}\,SE(\bar{X}_n)\big]$$

  or, in large samples,

$$\big[\bar{X}_n - 1.96\,SE(\bar{X}_n),\ \bar{X}_n + 1.96\,SE(\bar{X}_n)\big],$$

  where $t_{df,\,0.975}$ is the 0.975 quantile of the τ_{df} distribution.

5 Hypothesis testing (Lecture 2-3)

- Null and alternative hypotheses; steps of a test
- Type I error (its probability is the significance level, α)
- Type II error (its probability is 1 minus the power)

- One-sample t-test

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \tau_{n-1} \text{ under } H_0$$

- Two-sample t-test (large samples)

$$T = \frac{\bar{Y}_A - \bar{Y}_B}{\sqrt{S_A^2/n_A + S_B^2/n_B}} \overset{\text{approx.}}{\sim} N(0,1) \text{ under } H_0 \colon \mu_A = \mu_B$$

- p-values; critical values; one-sided versus two-sided tests
- Confidence intervals and tests: a 95% CI holds the values of μ_0 that a 5% two-sided test does not reject
- Practical versus statistical significance; power depends on effect size, variance and n

6 Bootstrap and clusters (Lecture 2-4)

- Nonparametric bootstrap: resample n observations with replacement, recompute the statistic, and use the standard deviation of the replicates as its standard error
- Parametric bootstrap
- Clustered data: G clusters of size m, intra-cluster correlation ρ

$$\operatorname{Var}(\bar{X}) = \frac{\sigma^2}{mG}\big(1 + (m-1)\rho\big)$$

- Cluster (block) bootstrap: resample whole clusters

7 What is regression? (Lecture 3)

- Regression as the conditional expectation function, $\mathbb{E}[Y \mid X = x]$
- Nonparametric regression: conditional means for discrete X; local averages (kernel regression) for continuous X
- Bandwidth and the bias-variance tradeoff

8 Simple regression (Lectures 4-1 to 4-3)

- OLS minimizes the sum of squared residuals (SSR)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\widehat{\operatorname{Cov}}(x,y)}{\widehat{\operatorname{Var}}(x)}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- Residuals sum to zero and are uncorrelated with x and $\hat{y}$; SST = SSE + SSR; R^2 = SSE/SST = 1 − SSR/SST
- Assumptions: (I) linearity, (II) random sampling, (III) variation in X, (IV) zero conditional mean, (V) homoskedasticity, (VI) normality
- Unbiasedness under I–IV; with a binary X, the slope is a difference in means
- Variance under I–V, and its estimate

$$\mathrm{Var}(\hat{\beta}_1 \mid X) = \frac{\sigma_u^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad \hat{\sigma}_u^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$$

- Gauss-Markov theorem: under I–V, OLS is the best linear unbiased estimator
- Inference under I–VI (exact) or in large samples (approximate)

$$T = \frac{\hat{\beta}_1 - \beta_1}{\widehat{SE}(\hat{\beta}_1)} \sim \tau_{n-2}, \qquad \hat{\beta}_1 \pm t_{n-2,\,0.975}\,\widehat{SE}(\hat{\beta}_1)$$

  (T is exactly τ_{n-2} under I–VI; without normality, approximately so, close to $N(0,1)$, in large samples)
- Consistency and asymptotic normality of OLS
- Interpretation: units, statistical significance, substantive significance, and no causal language without a design that justifies it
- Logs: level-level, log-level, level-log, log-log; a binary X with $\log Y$