Errata for "Statistics for Data Science" Course

Updated – December 20, 2017

1. Slide 27: If F_X is increasing and continuous, then $F_X^- = F_X^{-1}$.
2. Slide 29: Partial Converse Theorem → Let X be a random variable with continuous distribution F_X. Then, $F_X(X) \sim Unif(0,1)$.
3. Slide 39: The derivatives in the Jacobian matrix should be with respect to y_i and not x_i.
4. Slide 40: X and Y should also be independent.
5. Slide 40: The Jacobian matrix should be transposed; however this doesn't matter since we are interested in the determinant.
6. Slide 42: Properties of Expectation → $\mathbb{E}[h(X)]$ instead of $\mathbb{E}[h(x)]$.
7. Slide 48: Definition of Conditional Expectation → $\mathbb{E}[X \mid Y = y] = \sum_{x \in \mathcal{X}} x\mathbb{P}[X = x \mid Y = y]$, if X, Y are discrete.
8. Slide 49: If X is independent of Y, then $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$.
9. Slide 52: All instances of $\mathbb{R}^p$ should be $\mathbb{R}^d$.
10. Slide 54: $M_{X+Y} = M_X M_Y$ when X and Y are independent.
11. Slide 61: Lemma (Poisson and Multinomial) → $\mathbf{X} = (X_1, \ldots, X_k)^\top$.
12. Slide 68: The inequalities $H(X) \geq 0$ and $H(g(X)) \leq H(X)$ hold when X is discrete.
13. Slide 74:
    (a) "Natural" is from the mathematics point of view – usual parameter $\theta = \eta^{-1}(\phi)$ often different.
    (b) where $\eta : \mathbb{R}^k \to \mathbb{R}^k$ is a C^2 map such that $\phi = \eta(\theta)$.
14. Slide 88: since T and S are 1-1 functions of each other.
15. Slide 90:
    (a) without information loss on parameter of interest.
    (b) info can we throw away?
16. Slide 91: Notice that for any $\mathbf{y}$, $\mathbf{w}_{T(y)}$ is in the same level set as $\mathbf{y}$.
17. Slide 93: If the $\{T_j\}_{j=1}^k$ are non-trivial, the ratio $f(\mathbf{y})/f(\mathbf{z})$ will be constant with respect to $(\phi_1, \ldots, \phi_k)$ if and only if as $(\phi_1, \ldots, \phi_k)$ varies, the quantity below remains constant.
$$\sum_{j=1}^{k} \phi_j \left(\tau_j(y_1, \ldots, y_n) - \tau_j(z_1, \ldots, z_n)\right).$$
18. Slide 102: its form may be tedious to work with

19. Slide 103:
(a) $Q_n = n(1 - M_n)$.
(b) $\mathbb{P}[Q_n \leq y] = \mathbb{P}[M_n \geq 1 - y/n] = 1 - (1 - y/n)^n$.

20. Slide 104: $T(Y_1, \dots, Y_n) \xrightarrow{d} Z$.

21. Slide 105:
(a) $\mathbb{P}[|Y_n - Y| > \epsilon] \stackrel{n\to\infty}{\to} 0$.
(b) Example: Let $U_1, \dots, U_n \stackrel{iid}{\sim} \mathcal{U}[0,1]$ and $M_n = \max\{U_1, \dots, U_n\}$. Fix $\epsilon \in (0,1)$.

$$\mathbb{P}[|M_n - 1| > \epsilon] = \mathbb{P}[M_n > 1 + \epsilon] + \mathbb{P}[M_n < 1 - \epsilon] = 0 + (1 - \epsilon)^n \stackrel{n\to\infty}{\to} 0.$$

Also, $\mathbb{P}[|M_n - 1| > \epsilon] = 0$ if $\epsilon \geq 1$. Hence, $M_n \xrightarrow{p} 1$ as $n \to \infty$.

22. Slide 107: $\mathbb{P}[Y_n \leq y] = \mathbb{P}[Y_n \leq y, |Y_n - Y| \leq \epsilon] + \mathbb{P}[Y_n \leq y, |Y_n - Y| > \epsilon]$.

23. Slide 112: Theorem (Central Limit Theorem) Let $\{Y_n\}$ be an i.i.d. sequence with mean μ and variance $\sigma^2 < \infty$. Then,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu\right) \xrightarrow{d} N(0, \sigma^2).$$

24. Slide 118:
(a) Theorem (Multivariate Law of Large Numbers) Let $\{\mathbf{Y}_n\}$ be iid random vectors with $\mathbb{E}[\mathbf{Y}_k] = \mu$ and $\mathbb{E}||\mathbf{Y}_k|| < \infty$, for all $k \geq 1$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_k \xrightarrow{p} \mu.$$

(b) Theorem (Multivariate CLT) Σ should be replaced by Ω.

25. Slide 133: $A^*\widehat{\theta}$ should be replaced by $A^*(\theta)$.

26. Slide 135: $\mathbb{E}[\delta|T = t]$ in the first line of the proof should be replaced by $\mathbb{E}[\widehat{\theta}|T = t]$.

27. Slide 150: If both diagonal elements are positive, then it will be positive definite.

28. Slide 151: The likelihood is zero if $\theta < Y_{(n)}$.

29. Slide 155: The latter is minimized at θ and so $\Psi(\mathbf{u})$ is maximized at θ.

30. Slide 156:
(a) Does $\{\Psi_n(\mathbf{u}) \xrightarrow{p} \Psi(\mathbf{u})\ \forall\ \mathbf{u}$ with Ψ maximized uniquely at $\theta\} \dots$
(b) More general situations require stronger forms of convergence of $\Psi_n(\mathbf{u}) \to \Psi(\mathbf{u})$ plus additional regularity conditions.

31. Slide 159: In fact, the inverse function theorem tells us that the infinitely differentiable function $\Delta_\phi\gamma : \mathbb{R}^k \to \mathbb{R}^k$ must admit a continuously differentiable inverse map h locally ~~at $\phi \in \mathbb{R}^k$~~.

32. Slide 162: Note that this can be interpreted as

$$\widehat{\theta}_n \stackrel{d}{\approx} N\left(\theta, \frac{1}{n\mathcal{I}_1(\theta)}\right) \equiv N\left(\theta, \frac{1}{\mathcal{I}_n(\theta)}\right).$$

33. Slides 163 and 164: Why $\mathcal{I}_n(\theta)$? (... curvature)

34. Slide 173: $\tilde{\theta}_\alpha$ should be $\tilde{\mu}_a$ and $R(\tilde{\mu}, \mu)$ should be $MSE(\widehat{\mu}, \mu)$

35. Slide 180: all instances of $\overline{X}$ should be $\overline{Y}$

36. Slide 211: $I^{-1}(\theta)$ should be $\mathcal{I}_1^{-1}(\theta_0)$

37. Slide 219: $\mathbb{P}_{H_0}[T^2(\boldsymbol{Y}) \geq T(\boldsymbol{y})]$ should be $\mathbb{P}_{H_0}[T^2(\boldsymbol{Y}) \geq T^2(\boldsymbol{y})]$

38. Slide 265: the first term in the last equation should be $\mathbb{E}[\int_{\mathbb{R}} \widehat{f}_h^2(x)dx]$

39. Slide 266: the expectation in the first item should be $\mathbb{E}[\int_{\mathbb{R}} \widehat{f}_h^2(x)dx]$, and its estimator is $\int_{\mathbb{R}} \widehat{f}_h^2(x)dx$

40. Slide 266: the variables $\widehat{f}_{h,-i}(Y_i)$ are not independent

41. Slide 267: the LSCV should be $\int_{\mathbb{R}} \widehat{f}_h^2(x)dx - \frac{2}{n}\sum_{i=1}^n \widehat{f}_{h,-i}(Y_i)$

42. Slide 273: Y is a random output

43. Slide 295: the vectors $\{v_j\}$ in the lemma are not unique

44. Slide 296: $v_i \Sigma v_i^T$ should be $v_i^T \Omega v_i$

45. Slide 297: $X^T X$ should be XX^T

46. Slide 346: the correct definitions of the adjusted R^2 are

$$R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p} \quad R_{0a}^2 = 1 - (1 - R_0^2)\frac{n}{n-p}.$$

47. Slide 446: Z_j should be $(W_j - \mathbf{1}\overline{W}_j)/(sd(W_j)\sqrt{n})$ (divide rather than multiply by $\sqrt{n}$)

48. Slide 466: Y is a random output

49. Slide 470: $x_i \in \mathbb{R}^{p\times 1}$ so some transpositions are unnecessary:

$$\nabla_\beta \ell_n(\beta) = X_n^T Y - \sum_{i=1}^n x_i \gamma'(x_i^T \beta) = \sum_{i=1}^n x_i (Y_i - \mu_i) = X_n^T (Y - \mu)$$

50. Slide 472: the initialization should be $x_i^T \tilde{\beta} = (\gamma')^{-1}(Y_i)$