What to do today (Nov 30, 2020)?

1. *Introduction*
2. *Probability and Distribution (Chp 1-3)*
3. *Essential Topics in Mathematical Statistics*
4. **Further Topics, Selected from Chp 7-11**
   - **4.1 Nonparametric and Robust Statistics (Chp 10)**
     - *4.1.1 Location Models*
     - *4.1.2 Sample Median and the Sign Test*
     - *4.1.3 Signed-Rank and Mann-Whitney-Wilcoxon Tests*
     - **4.1.4 Measures of Association**
     - **4.1.5 Robust Concepts**
   - **4.2 Bayesian Procedures (Chp 11)**
     - **4.2.1 Prior and Posterior Distributions**
     - **4.2.2 Bayesian Point Estimation**
     - **4.2.3 Bayesian Interval Estimation and Testing**
     - *4.2.4 Additional Topics in Bayesian Statistics*

4.1.4 Measures of Association: Spearman's Rho

Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are iid observations on continuous bivariate (X, Y): **sample correlation coefficient** (Pearson correlation coefficient)

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

Definition. The statistic $r_S = \frac{\sum_{i=1}^{n}(R(X_i) - \frac{n+1}{2})(R(Y_i) - \frac{n+1}{2})}{n(n^2-1)/12}$ is called the **Spearman's rho**, where $R(X_i)$ and $R(Y_i)$ are the ranks of X_i and Y_i among $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively.

Proposition. If X and Y are independent, r_S is distribution free, symmetrically distributed with $E(r_S) = 0$ and $Var(r_S) = 1/(n-1)$. Plus, if $n >> 1$, $r_S/\sqrt{Var(r_S)} \sim N(0, 1)$ approximately.

Example 10.3 (cont'd) By

"cor.test(m1500,marathon,method="spearman")" in R:

$r_{S,obs} = 0.9053 \quad p - value = 2.021 - 06.$ $\Longrightarrow$ strong evidence against the hypothesis of the independence.

4.1.5 Robust Statistics Concepts (Chp10.9): sensitivity curve

Consider $X_1, \ldots, X_n$ iid observations on X, which follows the location model with the location parameter θ: $X = \theta + \epsilon$ with $\epsilon \sim f(\cdot)$. That is, $f_X(x) = f(x - \theta)$.

Definition. If $\hat{\theta} = \hat{\theta}(\mathbf{X}_n)$ is an estimator of θ with $\mathbf{X}_n = (X_1, \ldots, X_n)$ for $n \geq 1$, we call

$$S_n(x; \hat{\theta}) = \frac{\hat{\theta}(\mathbf{x}_{n+1}) - \hat{\theta}(\mathbf{x}_n)}{1/(n+1)}$$

the **sensitivity curve** of $\hat{\theta}$, where $\mathbf{x}_{n+1} = (\mathbf{x}_n, x)$.

- The sample mean $\bar{X}_n$ is an estimator of θ, the solution of $\sum_{i=1}^{n}(X_i - \theta) = 0$. It's sensitivity curve is $S_n(x; \bar{X}) = x - \bar{x}_n$. $\Longrightarrow$ the sample mean is quite sensitive to the size of outliers.
- The sample median $\hat{\theta}_n$ with odd n, for example, is $X_{([n+1]/2)}$. It's sensitivity curve $S_n(x; \hat{\theta})$ is bounded, not much sensitive to an outlier.

4.2 Bayesian Procedures (Chp 11)

Why to study Bayesian procedures?

- Two Statistical Schools:
  Frequentist Statistics vs Bayesian Statistics
  - Bayesian refers to Thomas Bayes (1702-1761).

- What do you know related to Bayesian statistics?
  - Subjective definition for probability
  - Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$$

  - EM Algorithm. ... revisit to the example ...
  - "Machine Learning", "Deep Learning", ...

4.2.1 Bayesian Procedures: Prior and Posterior Distributions

Consider rv $X \sim f(x|\theta)$.

- Frequentist viewpoint: $\theta \in \Omega$.
  That is, θ is unknown and taking value from Ω.

  By MLE, for example, to estm θ based on data $\mathbf{x}$ with

$$\widehat{\theta} = argmax_{\theta \in \Omega} L(\theta|\mathbf{x}).$$

- Bayesian viewpoint: $\Theta \sim h(\theta)$, a **prior distn**. The **posterior distn** given data $\mathbf{x}$ is

$$k(\theta|\mathbf{x}) = \left[\frac{L(\theta|\mathbf{x})}{\int L(\theta|\mathbf{x}) h(\theta) d\theta} \right] h(\theta) \propto L(\theta|\mathbf{x}) h(\theta).$$

4.2.2 Bayesian Procedures: Point Estimation

Bayes point estimator:

$$\hat{\theta}(\mathbf{x}) = argmin_{all\ \delta(\mathbf{x})}\left(E\left[\mathcal{L}(\Theta, \delta(\mathbf{x}))\right]\right)$$

with $E\left[\mathcal{L}(\Theta, \delta(\mathbf{x}))\right] = \int \mathcal{L}(\theta, \delta(\mathbf{x}))k(\theta|\mathbf{x})d\theta.$

- When the loss function is the squared difference $\mathcal{L}(\theta, \delta(\mathbf{x})) = (\delta(\mathbf{x}) - \theta)^2$,

$$\hat{\theta}(\mathbf{x}) = E(\Theta|\mathbf{x}) = \int \theta k(\theta|\mathbf{x})d\theta.$$

The posterior mean minimizes the mean squared error: $E\left[\left(\hat{\theta}(\mathbf{x}) - \theta\right)^2\right].$

4.2.3 Bayesian Procedures: Interval Estimation and Testing

- **Interval Estimation**. Interval $\big(u(\mathbf{x}), v(\mathbf{x})\big)$ is chosen such that

$$1-\alpha = P\big[u(\mathbf{x}) < \Theta < v(\mathbf{x})\big|\mathbf{x}\big] = \int_{u(\mathbf{x})}^{v(\mathbf{x})} k(\theta|\mathbf{x})d\theta.$$

  (*credible inverval*)

- **Testing Procedure**. To test on $H_0 : \theta \in \Omega_0$ vs $H_1 : \theta \in \Omega_1$:

  Accept H_0 if $P\big(\Theta \in \Omega_0|\mathbf{x}\big) \geq P\big(\Theta \in \Omega_1|\mathbf{x}\big)$; otherwise, reject H_0.

What will we study next?

1. *Introduction*

2. *Probability and Distribution (Chp 1-3)*

3. *Essential Topics in Mathematical Statistics (Chp 4-6)*

4. **Further Topics, Selected from Chp 7-11**
   - *4.1 Nonparametric and Robust Statistics (Chp 10)*
   - **4.2 Bayesian Procedures (Chp 11)**
     - *4.2.1 Prior and Posterior Distributions*
     - *4.2.2 Bayesian Point Estimation*
     - *4.2.3 Bayesian Interval Estimation and Testing*
     - **4.2.4 Additional Topics in Bayesian Statistics**

Plan for the remaining time of this term.