STAT 509 / ECON 580

HOMEWORK 5 (PART 1)

This homework will be due in quiz section in Gradescope on **Saturday**, November 17 at 9AM.

Note: I am still putting together a dataset for Qu.5. I will send an updated version that includes HW5 shortly.

Covariances

1. Let X_1 and X_2 be independent random variables, with means μ_1, μ_2, and variances σ_1^2 and σ_2^2 respectively. Further, let $S = (X_1 + X_2)/2$ and $T = (X_1 - X_2)/2$. Find:

   (a) $\mathrm{E}[S]$ and $\mathrm{E}[T]$;

   (b) $\mathrm{V}(S)$ and $\mathrm{V}(T)$;

   (c) $\mathrm{Cov}(S, T)$;

   (d) $\mathrm{Cov}(X_1, S)$, $\mathrm{Cov}(X_2, T)$.

Prediction

2. Suppose that X is a continuous random variable with support on $\mathbb{R}$. Suppose that the pdf for X is symmetric around a point t, so that $f(t - x) = f(t + x)$ for all x.

   (a) Find the median of X. *Hint: use the fact that the pdf integrates to* 1 *and then split the integral into two pieces.*

   (b) Find the mean of X. *Hint: use the fact that* $E[X] = t^*$ *if and only if* $E[X - t^*] = 0$. *Again split the integral; also see Midterm Qu.3(b)*

3. Consider a continuous random variable with pdf $f_X(x)$ and CDF $F_X(x)$.

   (a) Show that $\int_{-\infty}^{t} F_X(x)dx = \int_{-\infty}^{t} f(u)(t-u)du$.
   Hint: recall HW3 Qu.3

   (b) Similarly show that $\int_{t}^{\infty}(1 - F_X(x))dx = \int_{t}^{\infty} f(u)(u-t)du$.

   Define $h(t) \equiv E|X - t|$; the average absolute value of the prediction error, when using the constant t as a prediction (for all units).

   (c) Show that $h(t) = \int_{-\infty}^{t} F_X(x)dx + \int_{t}^{\infty}(1 - F_X(x))dx$.

   (d) Using your answer to (c) show that $h(t)$ is minimized when t is a median for X. Recall that m is a median for X if $F_X(m) = 0.5$.
   Hint: use the Fundamental Theorem of Calculus and your answer to (c).

4. Goldberger Question 6.2

5. *Data Question: In preparation - to follow - watch this space.*

Continued Over

Iterated Expectations

6. A population consists of two types, *humans* and *replicants*. The proportion of replicants is q. The height of each type approximately follow normal distributions. Let $N(\mu_H, \sigma_H^2)$ be the distribution of lengths for humans; let $N(\mu_R, \sigma_R^2)$ be the distribution of lengths for replicants.

   (a) Find the mean height of a randomly sampled subject in this population.

   (b) Find the variance of the distribution of height for subjects in this population.

7. Suppose that X and Y are continuous random variables, with support on $\mathbb{R}^2$. Suppose that two researchers, Thelma and Louise, wish to predict Y from X using a function of X.

   (a) Thelma wishes to use the function $g_T(X)$ that minimizes the average squared prediction $E[(Y - g_T(X))^2]$. What function will Thelma use? *You may justify your answer by quoting results from the Lecture.*

   (b) Louise, however, wishes to use the function $g_L(X)$ that minimizes the average absolute error $E[|Y - g_L(X)|]$. What function will Louise choose? Explain your answer. *Hint: Use the law of iterated expectations and Qu.3.*

   (c) Suppose that there is a function $r(x)$ such that the conditional density for Y given $X = x$ is symmetric around $r(x)$, so that for all x and y, $f(r(x) - y \mid x) = f(r(x) + y \mid x)$, what can we say about the functions $g_T(X)$ and $g_L(X)$ used by Thelma and Louise?

       Hint: Use Qu.2.

Continued Over

Bayesian Statistics

8. Suppose a medical test has the following characteristics:

$$\begin{aligned} Pr(\text{Test +ve} \mid \text{Patient Diseased}) &= 0.99 \\ Pr(\text{Test -ve} \mid \text{Patient Not Diseased}) &= 0.98 \end{aligned}$$

   (a) Find Pr(Test -ve | Patient Diseased) and Pr(Test +ve | Patient Not Diseased).

   Suppose that 1 in 5,000 people have this disease so

$$Pr(\text{Patient Diseased}) = 0.0002$$

   (b) Compute Pr(Test +ve). *Hint: Find Pr(Test +ve, Patient Diseased) and Pr(Test +ve, Patient Not Diseased).*

   (c) Use Bayes' rule to find Pr(Patient Diseased| Test +ve).

   (d) Give an intuitive explanation for the discrepancy between Pr(Patient Diseased | Test +ve) and Pr(Test +ve | Patient Diseased).