Problem Set 2 (Lectures 3 and 4)

Problem 1 (Optional): Let $S = 2$. Let $(x_1, x_2) \in \mathbb{R}^2$ and let $x_1^* \geq x_1$, $x_2^* \geq x_2$. Show that if F is the c.d.f. of an $\mathbb{R}^2$-valued random vector then:

$$F(x_1^*, x_2^*) - F(x_1^*, x_2) - F(x_1, x_2^*) + F(x_1, x_2) \geq 0$$

Problem 2 (Optional): Read the first three pages (290-292) of Cramér and Wold (1936) to see a sketch of the proof of the Cramér-wold Theorem (pretty much the same argument we make with moment generating functions). Then read their description of the "Moment Problem". Sketch the Moment Problem and the results concerning it.

COMMENT: The Cramér-wold device is something you will find in the proofs of multivariate Central Limit Theorems.

Problem 3: (The Best Linear Predictor, **30 Points**) Let X and Y be two real-valued random variables. Let

$$\mu \equiv \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \text{ and } \Sigma,$$

denote the mean and covariance of $(X, Y)'$. Both μ and Σ are assumed to be known. Suppose that we are interested in linearly "predicting" Y using the deviations of X from its mean. That is, we would like to predict Y using a function of the form

$$\alpha + \beta(X - \mu_x).$$

This is a purely probabilistic problem. There is no data, no sample, and no econometrics: just two-random variables. To assess the quality of the prediction, it seems natural to report the (squared)-distance between Y and its prediction (typically referred to as the *squared error*)

$$(Y - \alpha - \beta(X - \mu_x)).^2$$

The squared error is a random variable, which means that sometimes it will be big and sometimes it will be small. To summarize the performance of the linear predictor we can compute the average or mean squared-error:

$$\mathbb{E}[(Y - \alpha - \beta(X - \mu_x))^2].$$

The "best" linear predictor of Y in terms of $(X - \mu_x)$ is defined as the linear function $\alpha^* + \beta^*(X - \mu_x)$ that minimizes the mean squared-error; i.e., the function with parameters (α^*, β^*) such that:

$$(\alpha^*, \beta^*) \in \text{argmin}_{(\alpha, \beta)} \mathbb{E}[(Y - \alpha - \beta(X - \mu_x))^2].$$

1. (10/30 Points) Show first that

$$\alpha^* = \mu_y.$$

   This result implies that if X happens to be exactly equal to its mean, then the best linear predictor of Y is its expectation.

   HINT: Add and subtract μ_y and expand the square. You should not be doing a lot of algebra.

2. (10/30 Points) Show then that β^* solves the problem

$$\min_\beta \mathbb{E}[((Y - \mu_y) - \beta(X - \mu_x))^2],$$

   and therefore

$$\beta^* = Cov(X, Y)/V(X).$$

   HINT: The objective function is quadratic in β and the choice set for β is convex. You can see this by expanding the square and taking β outside the expectation operator. The problem is thus convex and the first order conditions are necessary and sufficient. One thing to notice in this problem is that $\frac{\partial}{\partial \beta} \mathbb{E}[((Y - \mu_y) - \beta(X - \mu_x))^2] = \mathbb{E}[\frac{\partial}{\partial \beta}((Y - \mu_y) - \beta(X - \mu_x))^2]$.

3. (10/30 Points) Finally, define the error of the best linear predictor as:

$$\epsilon = Y - \alpha^* - \beta^*(X - \mu).$$

   Show that $\mathbb{E}[\epsilon] = 0$ and $Cov(\epsilon, X) = 0$.

4. (OPTIONAL) Suppose now that instead of being real-valued, X is a random vector of dimension k. Suppose that we look for the coefficients α and $\beta = (\beta_1, \ldots, \beta_k)$ of the best linear predictor

$$\alpha + \beta_1(X_1 - \mu_1) + \beta_2(X_2 - \mu_2) + \ldots + \beta_k(X_k - \mu_k).$$

   Is it true that the coefficients of the best linear predictor are:

$$\alpha^* = \mu_y, \quad \beta = V(X)^{-1}Cov(X, Y)?$$

Problem 4: (Covariance equal to zero implies Independence?, **20 points**)

1. (10/20 Points) Show that if $(X, Y)'$ are bivariate normal random variables $Cov(X, Y) = 0$ implies independence. HINT: Use the definition of bivariate normal and any of the characterizations of independence we presented above.

2. (10/20 Points) Let $X \sim \mathcal{N}(0, 1)$. Let $Y = X^2$. Show that the $Cov(X, Y) = 0$. Are (X, Y) independent?

Problem 5: (Conditional distribution of a bivariate normal vector, **25 points**). Consider the same set-up as in the practice problem 3 where we introduced the best linear predictor. In this problem we will make the additional assumption that $(X, Y)' \sim \mathcal{N}_2((\mu_x, \mu_y)', \Sigma)$.

1. (20/25 points) What is the conditional distribution of $Y|X$?

2. (5/25 points) What is the relation between the best linear predictor of Y and the conditional expectation $E[Y|X]$ in the bivariate normal model?

HINT: There are two ways of doing this problem. One of them is long (and not very helpful): use the formula of conditional density and figure out the algebra. There is another shorter approach that I encourage you to use. Write:

$$Y = \alpha^* + \beta^*(X - \mu_x) + \epsilon, \quad \epsilon \equiv Y - \alpha^* - \beta^*(X - \mu_x),$$

and note that Y has been re-written as the sum of its best linear predictor and an "error" term, both of which are uncorrelated (by problem 3). Moreover, both the best linear predictor and the error term are normal random variables.

Problem 6: (Signal and Noise in a Bivariate Normal Model, **25 points**) Suppose that we have a random variable X that is generated additively by a "signal" θ and "noise" ϵ and:

$$X = \theta + \epsilon.$$

The signal and the noise are assumed to be independent normal random variables:

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad \theta \sim \mathcal{N}(\mu, \sigma_\theta^2).$$

Assume $\sigma_\epsilon^2, \sigma_\theta^2, \mu$ are all known. We would like to obtain the conditional distribution of the signal, given X, which is the "contaminated" or "noisy" measure of the signal.

1. Use problem 5 to show that $\theta|X$ is normal.

2. Show that $\mathbb{E}[\theta|X]$ is a linear combination between the mean of the signal μ and the value of the noisy measure x and that the weight on x is given by

$$\lambda = \sigma_\theta^2/(\sigma_\theta^2 + \sigma_\epsilon^2).$$

3. Show that the $\mathrm{Var}(\theta|X) = \sigma_\epsilon^2\lambda$.

References

CRAMÉR, H. AND H. WOLD (1936): "Some theorems on distribution functions," *Journal of the London Mathematical Society*, 1, 290–294.