Lecture 13

13.1 Minimal jointly sufficient statistics.

When it comes to jointly sufficient statistics $(T_1, \ldots, T_k)$ the total number of them (k) is clearly very important and we would like it to be small. If we don't care about k then we can always find some trivial examples of jointly sufficient statistics. For instance, the entire sample $X_1, \ldots, X_n$ is, obviously, always sufficient, but this choice is not interesting. Another trivial example is the order statistics $Y_1 \leq Y_2 \leq \ldots \leq Y_n$ which are simply the values $X_1, \ldots, X_n$ arranged in the increasing order, i.e.

$$Y_1 = \min(X_1, \ldots, X_n) \leq \ldots \leq Y_n = \max(X_1, \ldots, X_n).$$

$Y_1, \ldots, Y_n$ are jointly sufficient by factorization criterion, since

$$f(X_1, \ldots, X_n|\theta) = f(X_1|\theta) \times \ldots \times f(X_n|\theta) = f(Y_1|\theta) \times \ldots \times f(Y_n|\theta).$$

When we face different choices of jointly sufficient statistics, how to decide which one is better? The following definition seems natural.

Definition. (Minimal jointly sufficient statistics.) $(T_1, \ldots, T_k)$ are minimal jointly sufficient if given any other jointly sufficient statistics $(r_1, \ldots, r_m)$ we have,

$$T_1 = g_1(r_1, \ldots, r_m), \ldots, T_k = g_k(r_1, \ldots, r_m),$$

i.e. Ts can be expressed as functions of rs.

How to decide whether $(T_1, \ldots, T_k)$ is minimal? One possible way to do this is through the Maximum Likelihood Estimator as follows.

Suppose that the parameter set Θ is a subset of $\mathbb{R}^k$, i.e. for any $\theta \in \Theta$

$$\theta = (\theta_1, \ldots, \theta_k) \text{ where } \theta_i \in \mathbb{R}.$$

Suppose that given the sample $X_1, \ldots, X_n$ we can find the MLe of θ,

$$\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k).$$

The following simple fact will be useful.

Fact. *Given any jointly sufficient statistics* $(r_1, \ldots, r_m)$ *the MLE* $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ *is always a function of* $(r_1, \ldots, r_m)$.

To see this we recall that $\hat{\theta}$ is the maximizer of the likelihood which by factorization citerion can be represented as

$$f(x_1, \ldots, x_n|\theta) = u(x_1, \ldots, x_n)v(r_1, \ldots, r_m, \theta).$$

But maximizing this over θ is equivalent to maximizing $v(r_1, \ldots, r_m, \theta)$ over θ, and the solution of this maximization problem depends only on $(r_1, \ldots, r_m)$, i.e. $\hat{\theta} = \hat{\theta}(r_1, \ldots, r_m)$.

This simple fact implies that if MLE $\hat{\theta}$ is jointly sufficient statistics then $\hat{\theta}$ is minimal because $\hat{\theta} = \hat{\theta}(r_1, \ldots, r_m)$ for any jointly sufficient $(r_1, \ldots, r_m)$.

Example. If the sample $X_1, \ldots, X_n$ has uniform distribution $U[a, b]$, we showed before that

$$Y_1 = \min(X_1, \ldots, X_n) \text{ and } Y_n = \max(X_1, \ldots, X_n)$$

are the MLE of unknown parameters (a, b) and (Y_1, Y_n) are jointly sufficient based on factorization criterion. Therefore, (Y_1, Y_n) are minimal jointly sufficient.

Whenever we have minimal jointly sufficient statistics this yields one important consequence for constructing an estimate of the unkown parameter θ. Namely, if we measure the quality of an estimate via the average squared error loss function (as in the previous section) then Rao-Blackwell theorem tells us that we can improve any estimator by conditioning it on the sufficient statistics (this is also called projecting onto sufficient statistics). This means that any "good" estimate must depend on the data only through this minimal sufficient statistics, otherwise, we can always improve it. Let us give one example.

Example. If we consider a sample $X_1, \ldots, X_n$ from uniform distribution $U[0, \theta]$ then we showed before that

$$Y_n = \max(X_1, \ldots, X_n)$$

is the MLE of unknown parameter θ and also Y_n is sufficient by factorization criterion. Therefore, Y_n is minimal jointly sufficient. Therefore, any "good" estimate of θ should depend on the sample only through their maximum Y_n. If we recall the estimate of θ by *method of moments*

$$\hat{\theta} = 2\bar{X},$$

it is not a function of Y_n and, therefore, it can be improved.

Question. What is the distribution of the maximum Y_n?

End of material for Test 1. Problems on Test 1 will be similar to homework problems and covers up to Pset 4.

13.2 χ^2 distribution.

(Textbook, Section 7.2)

Consider a standard normal random variable $X \sim N(0,1)$. Let us compute the distribution of X^2. The cumulative distribution function (c.d.f.) of X^2 is given by

$$\mathbb{P}(X^2 \le x) = \mathbb{P}(-\sqrt{x} \le X \le \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

The p.d.f. is equal to the derivative $\frac{d}{dx}\mathbb{P}(X \le x)$ of c.d.f. and, hence, the density of X^2 is

$$\begin{aligned} f_{X^2}(x) &= \frac{d}{dx}\int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{x})^2}{2}} (\sqrt{x})' - \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{x})^2}{2}} (-\sqrt{x})' \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\frac{x}{2}} = \frac{1}{\sqrt{2\pi}} x^{\frac{1}{2}-1} e^{-\frac{x}{2}}. \end{aligned}$$

The probability density of X^2 looks like Gamma Distribution $\Gamma(\frac{1}{2}, \frac{1}{2})$. Recall that gamma distribution $\Gamma(\alpha, \beta)$ with parameters (α, β) has p.d.f.

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \text{ for } x \ge 0.$$

Consider independent random variables

$$X_1 \sim \Gamma(\alpha_1, \beta), \ldots, X_n \sim \Gamma(\alpha_n, \beta)$$

with gamma distributions that have the same parameter β, but $\alpha_1, \ldots, \alpha_n$ can be different. Question: what is the distribution of $X_1 + \ldots + X_n$?

First of all, if $X \sim \Gamma(\alpha, \beta)$ then the moment generating function of X can be computed as follows:

$$\begin{aligned} \mathbb{E}e^{tX} &= \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x} dx \\ &= \frac{\beta^\alpha}{(\beta-t)^\alpha} \underbrace{\int_0^\infty \frac{(\beta-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x} dx}. \end{aligned}$$

The function in the underbraced integral looks like a p.d.f. of gamma distribution $\Gamma(\alpha, \beta - t)$ and, therefore, it integrates to 1. We get,

$$\mathbb{E}e^{tX} = \Big(\frac{\beta}{\beta - t}\Big)^\alpha.$$

Moment generating function of the sum $\sum_{i=1}^n X_i$ is

$$\mathbb{E}e^{t\sum_{i=1}^n X_i} = \mathbb{E}\prod_{i=1}^n e^{tX_i} = \prod_{i=1}^n \mathbb{E}e^{tX_i} = \prod_{i=1}^n \Big(\frac{\beta}{\beta - t}\Big)^{\alpha_i} = \Big(\frac{\beta}{\beta - t}\Big)^{\sum \alpha_i}.$$

This means that:

$$\sum_{i=1}^n X_i \sim \Gamma\Big(\sum_{i=1}^n \alpha_i, \beta\Big).$$

Given i.i.d. $X_1, \cdots, X_n \sim N(0,1)$, the distribution of $X_1^2 + \ldots + X_n^2$ is $\Gamma(\frac{n}{2}, \frac{1}{2})$ since we showed above that $X_i^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.

Definition: χ_n^2 distribution with n degrees of freedom is the distribution of the sum $X_1^2 + \ldots + X_n^2$, where X_is are i.i.d. standard normal, which is also a gamma distribution $\Gamma(\frac{n}{2}, \frac{1}{2})$.