1. Probability Theory

GULP Winter 2015

March 9, 2015

1 Random Variables

We may think that the random variables (r.v.) are basically some random numbers. In graduate text the r.v. are **measurable functions** from the **sample space** to the real number space $\mathbb{R}^n$.

Example. Toss the coin twice. The sample space, which is all the result you can get, is HH,TT,HT,TH. Define the function as follows: $X = 1$ when you have two same result and $X = 0$ otherwise. Then $X(HT) = 0$, and $X(HH) = X(TT) = 1$.

As we define r.v. as functions, we can do integrations on it. The integration on the r.v. are done against the **probability measure**, which assigns weights to the sample space.

Like what we have in the language of analysis, if the r.v. has finite first moment, we call it is in L^1, or have a **finite mean**. Furthermore if it has finite second moment, we may conclude that it is in L^2, or has a **finite variance**. In probability space you always have finite lower moments when you have finite higher moments.

2 Densities and the MLE

It is not true anymore that only the continuous r.v. have the probability density. The density is used to describe the **possibility that some data appears**.

Example. The density of a continuous random variable is defined as

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

where F is the cdf for the r.v. If we do a bit of transformation:

$$f(x)\Delta x = F(x + \Delta x) - F(x) = \mathbb{P}(X \in (x, x + \Delta x))$$

it is roughly the probability that X falls into a small interval. Since the probability itself will go to zero when the interval shrinks to a number, we magnify the probability by dividing Δx so that it becomes the density function.

You may see that the pdf represents the comparative probability so that if we shift the pdf by a constant, it does not change anything if we do MLE or inference on that.

For discrete functions, the density function is naturally the probability mass function $\mathbb{P}(X = x_i)$, where $\{x_i\}$ are possible values that X can take. Clearly the density represents the possibility that data happens in this case.

Now we can define the density for other random variables.

Example Describe the density function for X such that $\mathbb{P}(X = 0) = 0.5$ and $\mathbb{P}(X = Y) = 0.5$ where Y follows an exponential distribution with parameter 1.

Example Describe the density function for X such that $\mathbb{P}(X = 0) = \theta$ and $\mathbb{P}(X = Y) = 0.5$ where Y follows an normal distribution with mean θ and variance 1.

When we have the idea of densities we can expand the idea of MLE.

Example(cont'd) Describe the likelihood function when we have n samples in the previous case.

$$L(\theta) = \prod_{i=1}^{n} \theta^{\mathbb{1}(X_i=0)} \phi_{\theta,1}(X_i)^{\mathbb{1}(X_i \neq 0)}$$

where

$$\phi_{\theta,1}(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{(x-\theta)^2}{2}}$$

The two parts of the likelihood function describe the relative and absolute possibility of observing the data in this case.

Example (Censoring) Suppose you have the survival data and assume the lifetime to follow exponential λ. Furthermore right censoring happens independently. Derive the MLE in this case.

Suppose the data is described as (X_i, C_i), where $C_i = 1$ when censoring happens, and 0 when no censoring happens. Then the 'likelihood' can be written as

$$L(\lambda) = \prod_{i=1}^{n} (\lambda e^{-\lambda x_i})^{1-c_i} (e^{-\lambda x_i})^{c_i}$$

3 Conditional Expectations

This is a topic which differs most from basic understanding to formal probability setting. When we consider the discrete case, it is well understood that $\mathbb{P}(Y|X = x)$ is well defined by the conditional probability formula. But the conditional probability formula can not properly define the conditional probability or the conditional expectation when X is continuous.

Then why conditional expectation are of the special interest? The conditional expectation $\mathbb{E}(Y|X)$ is **a functional of** X. Actually it is the best L^2 approximation of Y in all the measurable function of X.(You may find this sentence making no sense but it is OK now)

Then we have two property of the conditional expectations

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$$

and

$$\mathrm{Var}\,(X) = \mathbb{E}(\mathrm{Var}\,(X|Y)) + \mathrm{Var}\,(\mathbb{E}(X|Y))$$

By using this we can show some interesting theorems.