Statistical Inference I	**Prof. Chin-Tsang Chiang**
Lecture Notes 10	
October 15, 2015	**Scribe: Chi-Ning Chou**

1 Expectation

Expectation is simply a functional of the distribution. It maps a distribution to a certain real value to represent the behavior, shape, or other properties. Formally, we define the expectation of a random variable X as follow:

Definition 1 (expectation) *Let X be a r.v. and g be a measurable function. Then, the expectation of $g(X)$, which is also a r.v., is denoted as $\mathbb{E}[g(X)]$,i.e.,*

$$\mathbb{E}[g(X)] = \int_x g(x)dF_X(x)$$

Note that the expectation of $\mathbb{E}[g(X)]$ exists provided that $\mathbb{E}[|g(X)|] < \infty$.

Remark: If the distribution is not a mixture of both discrete and continuous distribution, then we can represent it as

- If X is discrete, $\mathbb{E}[g(X)] = \sum_x g(x)f_X(x)$.
- If X is continuous, $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)dF_X(x)$.

However, not all the distribution has expectation! Cauchy distribution is a beautiful example:
Example: (**Cauchy distribution has no mean**)
The pdf of Cauchy distribution is

$$f(x) = \frac{1}{\pi(1+x^2)}dx$$

With simple integration, we can check that $\mathbb{E}(|X|) = 2\int_0^\infty \frac{x}{\pi(1+x^2)}dx = \infty$. Thus, the expectation of Cauchy distribution does not exist. As a remark, Cauchy is a bell-shaped distribution with median 0. And actually, the cumulative distribution of Cauchy is the arc tangent function!

Property 2 *Let X be a r.v. and a,b,c be constants. Moreover, $g_1(X)$, $g_2(X)$ be any r.v. with expectation. Then,*

1. *(Preserve linear combination) $\mathbb{E}[]ag_1(X) + bg_2(X) + c] = a\mathbb{E}[g_1(X)] + b\mathbb{E}[g_2(X)] + c$.*
2. *(Preserve non-negativity) If $f(x) \geq 0,\ \forall x$, then $\mathbb{E}[g(X)]$.*
3. *(Preserve dominance) If $g_1(x) \geq g_2(x),\ \forall x$, then $\mathbb{E}[g_1(X)]gleq\mathbb{E}[g_2(X)]$.*
4. *(Existence of bounded r.v.) If $a \leq g(x) \leq b,\ \forall x$, then $a \leq \mathbb{E}[g(X)] \leq \mathbb{E}[g(X)]b$.*

Now, we turn to an useful and interesting application of expectation.

Example: (**The expectation of indicator function is probability**) Consider I_A to be an indicator function of a set $A \subseteq \mathbb{R}$, then

$$\mathbb{E}[I_A(X)] = P(A)$$

Moreover, we can regard the above equation as a **binary response**. That is, the indicator separate the space $\mathbb{R}$ into two parts: $\{x : x \in A\}$ and $\{x : x \notin A\}$ and the expectation is a functional to see the response of such partition.

For example, consider the following indicator function $I(X \leq x)$. We can see that $\mathbb{E}[I(X \leq x)] = F_X(x)$. And this representation gives us a broad way to describe the data. Suppose now we are concerning the probability $Pr[X = x|Z_1, Z_2, ..., Z_p]$, the most simply way is to use a general model to describe it, say

$$Pr[X = x|Z_1, Z_2, ..., Z_p] = G(x, \beta_1 Z_1 + \beta_2 Z_2 + ... + \beta_p Z_p)$$

As we choose to use the expectation representation: $\mathbb{E}[I(X \leq x)|Z_1, Z_2, ..., Z_p]$, the impact of Z_is can somehow depends on the value of x and become even more general. In other words, the linear parameter β_is can be depended on x. For example,

$$\begin{aligned} x_1 : \{x : X \leq x_1\} &\leftrightarrow \beta_{11} Z_1 + \beta_{12} Z_2 + ... + \beta_{1p} Z_p \\ x_2 : \{x : X \leq x_2\} &\leftrightarrow \beta_{21} Z_1 + \beta_{22} Z_2 + ... + \beta_{2p} Z_p \end{aligned}$$

With the above concept, we can simply show the inclusion-exclusion theorem with the help of indicator function and its expectation. First consider two facts:

- $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B$ and $\mathbf{1}_{A \cup B} = 1 - \mathbf{1}_{A^C \cap B^C}$
- $\mathbf{1}_{\cup_i A_i} = 1 - \prod_i (1 - \mathbf{1}_{A_i})$

Now, we can derive the inclusion-exclusion theorem:

$$\begin{aligned} P(\cup A_i) &= 1 - \mathbb{E}[\prod_i (1 - \mathbf{1}_{A_i})] \\ &= 1 - \mathbb{E}[1 - \sum_i \mathbf{1}_{A_i} + \sum_{i,j} \mathbf{1}_{A_i} \mathbf{1}_{A_j} - ... + (-1)^k \sum_{i_1,...,i_k} \mathbf{1}_{A_{i_1}} \cdots \mathbf{1}_{A_{i_k}} \pm ... \pm \mathbf{1}_{\cap A_i}] \\ &= \sum_i P(A_i) - \sum_{i,j} P(A_i \cap A_j) + ... + (-1)^{k-1} \sum_{i_1,...,i_k} P(A_{i_1} \cap \cdots \cap A_{i_k}) \pm ... \pm P(\cup_i A_i) \end{aligned}$$