

## Lecture Notes 17

November 21, 2015

Scribe: Wei-Chang Lee, Chi-Ning Chou

## 0.1 Rose: Poisson process and Renewal process

Nonhomogeneous Poisson process deals with the situation that the rate is not fixed. Instead, it is a function of time, that is,  $\lambda(t)$ . Professor Chiang showed three examples about nonhomogeneous Poisson process as follow:

## 0.1.1 How to generate nonhomogeneous Poisson process

Suppose we want to generate  $N^*(t) \sim \text{Poisson}(\lambda(t))$ , what can we do? By definition, we have

$$P[N^*(t+h) - N^*(t) = 1] = \lambda(t)h + o(h)$$

Let  $m(t) = \int_0^t \lambda(u)du$ , then we can show that

$$P[N^*(t+s) - N^*(t) = n] = \frac{(m(t+s) - m(t))^n e^{-(m(t+s) - m(t))}}{n!}$$

Now, we consider a homogeneous Poisson process  $N(t) \sim \text{Poisson}(\lambda)$ . Similarly, we have

$$P[N(t+s) - N(t) = n] = \frac{\lambda^n e^{-\lambda}}{n!}$$

By definition, the rate of nonhomogeneous Poisson process varies by times. Intuitively, we can think of this phenomenon in another way: the time unit is scaling time by time. For example, in the small time interval  $[t, t+h]$  where the rate almost being constant. We have

$$P[N(t+h) - N(t) = 1] = h\lambda + o(h)$$

and

$$\begin{aligned} P[N^*(t+h) - N^*(t) = 1] &= h\lambda(t) + o(h) \\ &= (h\frac{\lambda(t)}{\lambda})\lambda + o(h\frac{\lambda(t)}{\lambda}) \\ &= P[N(t+h\frac{\lambda(t)}{\lambda}) - N(t) = 1] \end{aligned}$$

To sum up, we will have

$$P[N^*(t+s) - N^*(t) = n] = P[N(\frac{m(t+s)}{\lambda}) - N(\frac{m(t)}{\lambda}) = n]$$

Namely, we can generate a nonhomogeneous Poisson process  $N^*(t) \sim \text{Poisson}(\lambda(t))$  with a standard Poisson process  $N(t) \sim \text{Poisson}(\lambda)$  in the following way

$$N^*(t) = N(\frac{m(t)}{\lambda})$$

### 0.1.2 Latent effects will be ruled out as fixing observation time

## 0.2 Location-scale family

**Definition 1 (location-scale family)** We say a family of distribution with standard pdf  $f$  is

$$\mathcal{F} = \{g : \mu \in \mathbb{R}, \sigma > 0, \forall x \in \mathbb{R}, \sigma g(\frac{x - \mu}{\sigma}) = f(x)\}$$

We can also characterize a location-scale family in another way presented in the following theorem.

**Theorem 1** Let  $f$  be a pdf and  $\mu \in \mathbb{R}, \sigma > 0$ , then

$$X \sim \frac{1}{\sigma} f(\frac{x - \mu}{\sigma}) \Leftrightarrow \exists Z \in f(z) \text{ s.t. } X = \sigma Z + \mu$$

**Proof:**

( $\Leftarrow$ ) Define  $g(z) = \sigma z + \mu$ . As  $g(\cdot)$  is monotone and measurable, we can conclude that  $X = g(Z)$  is a random variable with pdf  $f_X(x) = \frac{1}{\sigma} f(\frac{x - \mu}{\sigma})$

( $\Rightarrow$ ) Define  $g(x) = \frac{x - \mu}{\sigma}$ . As  $g(\cdot)$  is monotone and measurable, we can conclude that  $Z = g(X)$  is a random variable with pdf  $f_Z(z) = f(z)$ . ■

With Theorem 1, we can see that when we know the behavior of  $Z$ , we can use some linear transformation to infer the behavior of random variables that follow the distribution in the same location-scale family. For example, we have

- $\mathbb{E}[X] = \sigma \mathbb{E}[Z] + \mu$
- $Var[X] = \sigma^2 Var[Z]$

Moreover, location-scale family is *stochastically increasing*. Recall that

**Definition 2 (stochastically increasing)** A family of distribution  $\{f(x|\theta) : \theta \in \Theta\}$  is stochastically increasing in  $\theta$  is  $\forall \theta_1 > \theta_2$ ,

$$F(x|\theta_1) < F(x|\theta_2), \forall x$$

In other words,

$$Tail \text{ probability of } \theta_1 > Tail \text{ probability of } \theta_2$$

**Remark:** Geometrically, the cdf of  $\theta_2$  is strictly above that of  $\theta_1$ .

## 0.3 Probability inequalities

### 0.3.1 Markov inequality

When a random variable  $X$  is **nonnegative**, Markov inequality provides an upper bound for the tail probability of  $X$  as follow

$$P[X \geq r] \leq \frac{\mathbb{E}[X]}{r}$$

**Proof:** The proof is one line:

$$\mathbb{E}[X] = \int_0^\infty x dF_X(x) \geq \int_r^\infty x dF_X(x) \geq r P[X \geq r]$$

■

### Intuition (Markov inequality)

The idea of Markov inequality and other related inequalities is quite elegant. We separate the integral in two parts, then upper bound one of them by the nonnegativity of probability and upper the other by the monotonicity of the support.

### 0.3.2 Chebyshev inequality

We can generalize Markov inequality by transforming random variables into nonnegative random variables. And this is exactly the central idea of Chebyshev inequality.

**Theorem 2 (Chebyshev inequality)** *Let  $X$  be a random variable and  $g(x)$  be a nonnegative measurable function. Then,  $\forall r > 0$ , we have*

$$P[g(X) \geq r] \leq \frac{\mathbb{E}[g(X)]}{r}$$

However, we can see that the probability inequalities in this fashion are very loose, just like Boole's inequality. As a result, they will definitely overestimate. Thus, the most important applications of these inequalities are not bounding error. Instead, they have excellent performance on bounding *error rate*.