Lecture 25

Andrei Antonenko

March 30, 2005

1 Poisson Approximation to Binomial Distribution: Reminder

In the previous section we saw that the Binomial distribution can be approximated by Poisson distribution. Precisely, if n is large, and p is small, we have

$$\frac{n!}{(n-k)!k!}p^k(1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k!}, \tag{1}$$

where

$$\lambda = np.$$

This approximation is normally used when $n \geq 100$ and $p \leq 0.01$.

2 Poisson Processes

Let's consider the sequence of traffic accidents in Manhattan. In order to describe this process, we can break the time line into minutes, and reduce the problem to the Bernoulli process – accidents happens during each minute with some small probability p or it does not happen during this minute with respective probability $1-p$. The problem with this approach is that there might be more than one accident per minute in the City, and therefore, we will not be able to describe the total number of accidents in the City by the Bernoulli process. We can break the time line into smaller periods – seconds, milliseconds, etc. – such periods that we will believe that it is impossible for two accidents happen during this period. Instead of that we will try to derive a different approach.

Below we will talk instead of accidents about arrivals. That is a general term used in the theory of random processes for events, happening at different moments of time.

Let's introduce probability $P(k,\tau)$ of k arrivals during the time interval of length τ. The properties we will be assuming about $P(k,\tau)$ are the following:

(1) **[Homogeneity.]** $P(k,\tau)$ is the same for all interval of the length τ. In our example it means for example that the probability of having, say, 5 accidents between 8pm and 9pm is equal to the probability of having 5 accidents between 10am and 11am. Stating differently, the arrivals are equally likely at all times.

(2) **[Independence.]** Number of the arrivals during the interval is independent of the number of the history outside the interval. In our example, it means, that the number of accidents between 1pm and 2pm is independent on the number of accidents, which happened before 1pm. Stating differently, information about the arrivals outside the interval is irrelevant for arrivals inside the interval.

(3) **[Small interval probabilities.]** We will assume that

$$P(0,\tau) = 1 - \lambda\tau + o(\tau) \tag{2}$$

$$P(1,\tau) = \lambda\tau + o_1(\tau) \tag{3}$$

as $\tau \to 0$. Here, $o(\tau)$ and $o_1(\tau)$ are negligible in comparison to τ. Thus, for small τ, the probability of a single arrival is $\lambda\tau$ and the probability of no arrivals is $1 - \lambda\tau$, when the probability of 2 or more arrivals is negligible as τ gets smaller.

Now using these properties, we will derive the formula for $P(k,\tau)$.

Let's take a period of time of length τ, and divide it into periods of length δ, such that in total we will have τ/δ periods. Probability of 2 or more arrivals during each period is negligible by property (3). Moreover, all the periods are independent by property (2). Probability of 1 arrival during each period is $\lambda\delta$, and the probability of no arrivals is $1 - \lambda\delta$. Therefore, we see that $P(k,\tau)$ is a binomial probability of k successes in $n = \tau/\delta$ trials with probability of success $\lambda\delta$. As δ decreases, it can be approximated by Poisson distribution with parameter $\mu = \lambda\delta \cdot \frac{\tau}{\delta} = \lambda\tau$, and we obtain:

$$P(k,\tau) = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}, \tag{4}$$

where λ is a rate of the arrivals (number of arrivals per unit time). From here we can see that the expected number of arrivals N_τ during the period of time τ is the mean of Poisson distribution with parameter $\lambda\tau$, and the variance is the variance of Poisson distribution with parameter $\lambda\tau$:

$$\mathbf{E}[N_\tau] = \lambda\tau; \quad \mathbf{var}(N_\tau) = \lambda\tau. \tag{5}$$

Example 2.1. Assume that emails to your account are arriving with the rate $\lambda = 0.2$ messages per hour (1 email per 5 hours). Assume you check your mail every hour. What is the probability that you will find 0 messages in your mailbox? 1 message in your mailbox?

Using the formulae above, we have:

$$P(0,1) = e^{-0.2\cdot 1}\frac{(0.2\cdot 1)^0}{0!} = e^{-0.2} = 0.819;$$

$$P(1,1) = e^{-0.2\cdot 1}\frac{(0.2\cdot 1)^1}{1!} = 0.2\cdot e^{-0.2} = 0.164.$$

Now assume that you didn't check your messages for the whole day. What is the probability that there will be no new mail in your inbox?

$$P(0,24) = e^{-0.2\cdot 24} = 0.008294.$$