Indian Institute of Technology Madras

Department of Data Science and Artificial Intelligence

DA5000: Mathematical Foundations of Data Science

Tutorial V - Solutions

# Solutions

1. To prove that the binomial distribution converges to the Poisson distribution as $n \to \infty$ and $p \to 0$, subject to the constraint $np = \lambda$, we proceed as follows.

   Step 1: Binomial PMF

   The probability mass function (PMF) of a binomial distribution $X \sim \text{Binomial}(n, p)$ is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots$$

   We want to show that this converges to the PMF of a Poisson distribution as $n \to \infty$ and $p \to 0$, with $np = \lambda$.

   Step 2: Substituting $p = \frac{\lambda}{n}$

   Since $p = \frac{\lambda}{n}$, substitute this into the binomial PMF:

$$P(X = k) = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

   Step 3: Simplifying the factorial expression

   The binomial coefficient can be expressed as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

   Thus, the binomial PMF becomes:

$$P(X = k) = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

   Now, we simplify the terms individually as $n \to \infty$ and $p \to 0$.

   Step 4: Evaluate the terms as $n \to \infty$

   1. The binomial coefficient:

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} = 1 \cdot \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)$$

   As $n \to \infty$, each term $\left(1 - \frac{i}{n}\right)$ for $i = 1, 2, \ldots, k-1$ approaches 1. Therefore:

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \to 1$$

   2. The term $\left(1 - \frac{\lambda}{n}\right)^{n-k}$:

We use the fact that $\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$ as $n \to \infty$. Specifically:

$$\left(1-\frac{\lambda}{n}\right)^{n-k} = \left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k}$$

As $n \to \infty$, $\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$, and $\left(1-\frac{\lambda}{n}\right)^{-k} \to 1$, because $\left(1-\frac{\lambda}{n}\right) \approx 1$ for large $n$.

Thus, $\left(1-\frac{\lambda}{n}\right)^{n-k} \to e^{-\lambda}$.

Step 5: Putting it all together

Now, combining these results:

$$P(X=k) = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

As $n \to \infty$, this becomes:

$$P(X=k) \to \frac{\lambda^k}{k!}e^{-\lambda}$$

This is the PMF of a Poisson distribution with parameter $\lambda$. Thus, we have shown that as $n \to \infty$ and $p \to 0$, with $np = \lambda$, the binomial distribution converges to the Poisson distribution.

2. Given, 1.5 % of all mobile phones manufactured are defective. This implies that probability of defective units is $\frac{1.5}{100} = 0.015$. We are given that n=700. As n is large here and p is small, we can approximate it as a poisson process. $\lambda = np = 0.015 * 700 = 10.5$. We know P(X=x) is given by the Poisson Distribution Formula as $\frac{e^{-\lambda}\lambda^x}{x!}$. $P(X<3) = P(X=0) + P(X=1) + P(X=2)$. Substituting the values, we get $P(X<3) = \frac{e^{-10.5}10.5^0}{0!} + \frac{e^{-10.5}10.5^1}{1!} + \frac{e^{-10.5}10.5^2}{2!} = 1.8346 * 10^{-3}$. Therefore the final answer is $P(X<3) = 1.8346 * 10^{-3}$.

   If the incidence rate is reduced to 0.5 %, the new mean is given by $\lambda = np = 0.005 * 700 = 3.5$. Applying the same formula for P, we get $P(X<3) = \frac{e^{-3.5}3.5^0}{0!} + \frac{e^{-3.5}3.5^1}{1!} + \frac{e^{-3.5}3.5^2}{2!} = 0.32$. Therefore, the final answer is $P(X<3) = 0.32$

3. **1. Exponential Distribution (for Random Variable $X$)**

### Expectation

The expectation $\mathbb{E}[X]$ is given by:

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x}\,dx$$

Using integration by parts: - Let $u = x$ and $dv = \lambda e^{-\lambda x}\,dx$ - Then, $du = dx$ and $v = -e^{-\lambda x}$

Applying integration by parts:

$$\mathbb{E}[X] = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\,dx$$

The first term evaluates to 0, and the second integral gives:

$$\int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}$$

Thus,

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

### CDF

The CDF $F_X(x)$ is:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t}\, dt$$

Evaluating this integral:

$$F_X(x) = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}$$

For $x < 0$, $F_X(x) = 0$. So the CDF is:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

## 2. Bernoulli Distribution (for Random Variable $Y$)

### Expectation

The expectation $\mathbb{E}[Y]$ is given by:

$$\mathbb{E}[Y] = 0 \cdot (1-p) + 1 \cdot p = p$$

Thus, the expectation has a closed-form solution:

$$\mathbb{E}[Y] = p$$

### CDF

The CDF $F_Y(k)$ for the Bernoulli distribution is:

$$F_Y(k) = P(Y \leq k) = \begin{cases} 0 & \text{for } k < 0 \\ 1-p & \text{for } 0 \leq k < 1 \\ 1 & \text{for } k \geq 1 \end{cases}$$

This expression is also in closed form.

## Weibull Distribution

The Weibull distribution is defined by its probability density function (PDF):

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where $\lambda > 0$ is the scale parameter and $k > 0$ is the shape parameter.

### Expectation

The expectation $\mathbb{E}[X]$ of the Weibull distribution does not have a simple closed-form expression for arbitrary $k$. It is given by:

$$\mathbb{E}[X] = \lambda \Gamma\left(1 + \frac{1}{k}\right)$$

where $\Gamma$ is the gamma function. The gamma function does not have a closed-form solution for non-integer values and requires numerical methods for evaluation.

**CDF**

The CDF of the Weibull distribution is:

$$F(x; \lambda, k) = 1 - e^{-(x/\lambda)^k} \quad \text{for } x \geq 0$$

While the CDF itself is in closed form, the lack of a closed-form solution for the expectation when $k$ is arbitrary indicates that the Weibull distribution generally does not have a closed-form solution for its mean.

4. The mean of a uniform distribution over $[5, 15]$ is:

$$\mu = \frac{5 + 15}{2} = 10$$

The value is closer to the mean if it lies within half the distance from the mean to either endpoint. The endpoints are at distances of:

$$|10 - 5| = 5$$

From mean to upper endpoint:

$$|15 - 10| = 5$$

Thus, values closer to the mean lie within [7.5, 12.5].

The probability is given by the length of this interval divided by the total interval length:

$$P(7.5 \leq X \leq 12.5) = \frac{12.5 - 7.5}{15 - 5} = \frac{5}{10} = 0.5$$

5. The memoryless property states that for all $s, t \geq 0$:

$$P(X > s + t \mid X > s) = P(X > t).$$

Show that the exponential distribution satisfies this property.

**Solution:** The exponential distribution with parameter $\lambda > 0$ has the probability density function (PDF):

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Its cumulative distribution function (CDF) is:

$$F_X(x) = P(X \leq x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

The survival function, which gives the probability that the random variable exceeds a certain value, is:

$$P(X > x) = 1 - F_X(x) = e^{-\lambda x}, \quad x \geq 0.$$

We want to show that the exponential distribution satisfies the memoryless property:

$$P(X > s + t \mid X > s) = P(X > t).$$

Using the definition of conditional probability, we have:

$$P(X > s + t \mid X > s) = \frac{P(X > s + t \cap X > s)}{P(X > s)}.$$

Since $X > s+t$ implies $X > s$, the intersection $P(X > s+t \cap X > s)$ is simply $P(X > s+t)$. Therefore, we can rewrite the equation as:

$$P(X > s+t \mid X > s) = \frac{P(X > s+t)}{P(X > s)}.$$

Using the survival function $P(X > x) = e^{-\lambda x}$, we get:

$$P(X > s+t \mid X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}.$$

But $e^{-\lambda t} = P(X > t)$, which proves that the exponential distribution satisfies the memoryless property:

$$P(X > s+t \mid X > s) = P(X > t).$$