Engineering Mathematics and Statistics (B39AX)
Fall 2023

Tutorial 4

Problem A. The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space. It is parameterized by $\lambda \geq 0$, and its probability mass function is

$$p(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k\,!}\,,$$

for $k = 0, 1, 2, \ldots$.

(a) Show that the expected value of a random variable with Poisson distribution is λ.

Using the definition of expected value for discrete random variables, if X has a Poisson distribution, then

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} k \cdot p_X(k\,;\,\lambda) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k\,!} = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k\,!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k\,!} && \text{(first term is 0)} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{z=0}^{\infty} \frac{\lambda^z}{z\,!} && \text{(change of variable: } z = k-1) \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda\,.
\end{aligned}$$

(b) The number of people that arrive at a shop within a 10 minute period is Poisson distributed with mean $\lambda = 2$. What is the probability nobody arrives within a 15 minute period?

Let X be the number of people that arrive within a 10 minute period. Then, $X \sim \text{Poisson}(\lambda_X)$, with $\lambda_X = 2$. Let Y be the number of people that arrive within a 15 minute period. We have $Y = \frac{15}{10}X = \frac{3}{2}X$, and Y is Poisson distributed with mean λ_Y. To find out λ_Y, use (a) and the linearity of the expected value:

$$\lambda_Y = \mathbb{E}[Y] = \mathbb{E}\Big[\frac{3}{2}X\Big] = \frac{3}{2}\mathbb{E}[X] = \frac{3}{2}\lambda_X = 3\,.$$

The probability that nobody arrives within a 15 minute period is then

$$\mathbb{P}(Y = 0) = p_Y(0;3) = \frac{3^0 \cdot e^{-3}}{0\,!} = e^{-3} \simeq= 0.0498\,,$$

that is, 4.98%.

Problem B. Markov's inequality states that any nonnegative random variable $X \geq 0$ satisfies[1]

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}\,,$$

for any $c > 0$. Use it to show Chebyshev's inequality: for any random variable X and any $c > 0$,

$$\mathbb{P}\Big(|X - \mathbb{E}[X]| \geq c\Big) \leq \frac{\mathrm{Var}(X)}{c^2}\,.$$

Hint: Apply Markov's inequality to the random variable $Y = \big(X - \mathbb{E}[X]\big)^2$.

$$\mathbb{P}\Big(|X - \mathbb{E}[X]| \geq c\Big) = \mathbb{P}\Big(\big(X - \mathbb{E}[X]\big)^2 \geq c^2\Big) = \mathbb{P}(Y \geq c^2) \leq \frac{\mathbb{E}[Y]}{c^2} = \frac{\mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big]}{c^2} = \frac{\mathrm{Var}(X)}{c^2}\,.$$

The inequality is due to Markov's inequality.

Problem C. We toss a fair coin 40 times and want to compute the probability that we get at least 30 heads. Use two methods:

a) The central limit theorem.

We first identify the set of random variables

$$X_i = \text{“Obtain heads in the } i\text{th toss”},$$

for $i = 1, \ldots, n = 40$. Each X_i is a Bernoulli random variable with parameter $p = 1/2$. Therefore, $\mathbb{E}[X_i] = p = 1/2$ and $\mathrm{Var}(X_i) = p(1-p) = 1/4$. We are asked

$$\mathbb{P}\Big(X_1 + \cdots + X_n \geq 30\Big)\,.$$

To use the central limit theorem, we first need to compute the expected value and variance of $X_1 + \cdots + X_n$:

$$\mathbb{E}\Big[X_1 + \cdots + X_n\Big] = \underbrace{\mathbb{E}[X_1]}_{1/2} + \cdots + \underbrace{\mathbb{E}[X_n]}_{1/2} = \frac{n}{2} = 20$$

$$\mathrm{Var}\Big(X_1 + \cdots + X_n\Big) = \underbrace{\mathrm{Var}(X_1)}_{1/4} + \cdots + \underbrace{\mathrm{Var}(X_n)}_{1/4} = \frac{n}{4} = 10\,.$$

Recall that the standard deviation is $\sqrt{\mathrm{Var}(X_1 + \cdots + X_n)} = \sqrt{10} \simeq 3.162$. Using Z to represent a normal standard variable $\mathcal{N}(0,\,1)$, we then have

$$\mathbb{P}\Big(X_1 + \cdots + X_n \geq 30\Big) = \mathbb{P}\Big(\frac{X_1 + \cdots + X_n - 20}{3.162} \geq \frac{30 - 20}{3.162}\Big)$$

[1]This can be shown by the following argument (for continuous random variables)

$$\mathbb{E}[X] = \int_0^{+\infty} x \cdot f_X(x)\,dx \geq \int_c^{+\infty} x \cdot f_X(x)\,dx \geq c\int_c^{+\infty} f_X(x)\,dx = c \cdot \mathbb{P}(X \geq c)\,,$$

where the first inequality uses the fact that $x \cdot f_X(x) \geq 0$ for $x \geq 0$, and the second one the fact that $x \cdot f_X(x) \geq c \cdot f_X(x)$ for all $x \geq c$.

$$
\begin{aligned}
&\simeq \mathbb{P}(Z \geq 3.16) \\
&= 1 - \mathbb{P}(Z \leq 3.16) \\
&= 1 - \Phi(3.16) \\
&\simeq 1 - 0.9992 \\
&= 0.0008
\end{aligned}
$$

where the approximation in the second step is due to the application of the central limit theorem. This means the probability is 0.08%.

b) Chebyshev's inequality (from the previous problem).

Using $S_n := X_1 + \cdots + X_n$ and Chebyshev's inequality, we have

$$\mathbb{P}(S_n \geq 30) = \mathbb{P}(S_n - \underbrace{20}_{\mathbb{E}[S_n]} \geq 10) \leq \mathbb{P}(|S_n - 20| \geq 10) \leq \frac{\mathrm{Var}(S_n)}{10^2} = 0.1\,.$$

In other words, the probability is 10%. Notice that Chebyshev's inequality provides a loose upper bound in this case.