EEF 271E
Probability and Statistics

Week #7

EXAMPLE

Find the covariance of the random variables whose joint probability density is given by

$$f(x, y) = \begin{cases} 2 & \text{for } x > 0, y > 0, x + y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Solution Evaluating the necessary integrals, we get

$$\mu_X = \int_0^1 \int_0^{1-x} 2x\,dy\,dx = \frac{1}{3}$$

$$\mu_Y = \int_0^1 \int_0^{1-x} 2y\,dy\,dx = \frac{1}{3}$$

$$\sigma'_{1,1} = \int_0^1 \int_0^{1-x} 2xy\,dy\,dx = \frac{1}{12}$$

$$\sigma_{XY} = \frac{1}{12} - \frac{1}{3} \cdot \frac{1}{3} = -\frac{1}{36}$$

THEOREM If X and Y are independent, then $E(XY) = E(X) \cdot E(Y)$ and $\sigma_{XY} = 0$.

Proof. For the discrete case we have, by definition,

$$E(XY) = \sum_x \sum_y xy \cdot f(x, y)$$

Since X and Y are independent, we can write $f(x, y) = g(x) \cdot h(y)$, where $g(x)$ and $h(y)$ are the values of the marginal distributions of X and Y, and we get

$$E(XY) = \sum_x \sum_y xy \cdot g(x)h(y)$$

$$= \left[\sum_x x \cdot g(x)\right]\left[\sum_y y \cdot h(y)\right] = E(X) \cdot E(Y)$$

Hence,

$$\begin{aligned}
\sigma_{XY} &= \mu'_{1,1} - \mu_X \mu_Y \\
&= E(X) \cdot E(Y) - E(X) \cdot E(Y) \\
&= 0
\end{aligned}$$

It is of interest to note that the independence of two random variables implies a zero covariance, but a zero covariance does not necessarily imply their independence.

EXAMPLE

If the joint probability distribution of X and Y is given by

		x = −1	x = 0	x = 1	
	-1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{2}{3}$
y	0	0	0	0	0
	1	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$
		$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

Solution Using the probabilities shown in the margins, we get

$$\mu_X = (-1)\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = 0$$

$$\mu_Y = (-1)\cdot\frac{2}{3} + 0\cdot 0 + 1\cdot\frac{1}{3} = -\frac{1}{3}$$

and

$$\begin{aligned}\mu'_{1,1} &= (-1)(-1)\cdot\frac{1}{6} + 0(-1)\cdot\frac{1}{3} + 1(-1)\cdot\frac{1}{6} \\ &\quad + (-1)1\cdot\frac{1}{6} + 1\cdot 1\cdot\frac{1}{6} \\ &= 0\end{aligned}$$

Thus, $\sigma_{XY} = 0 - 0(-\frac{1}{3}) = 0$,
the covariance is zero, but the two random variables are not independent.

THEOREM If $X_1, X_2, \ldots, X_n$ are independent, then

$$E(X_1 X_2 \cdot \ldots \cdot X_n) = E(X_1) \cdot E(X_2) \cdot \ldots \cdot E(X_n)$$

CONDITIONAL EXPECTATIONS

Conditional expectations of random variables are likewise defined in terms of their conditional distributions.

DEFINITION

If X is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at x, the **conditional expectation** of $u(X)$ given $Y = y$ is

$$E[u(X)|y] = \sum_{x} u(x) \cdot f(x|y)$$

Correspondingly, if X is a continuous random variable and $f(x|y)$ is the value of the conditional probability density of X given $Y = y$ at x, the **conditional expectation** of $u(X)$ given $Y = y$ is

$$E[u(X)|y] = \int_{-\infty}^{\infty} u(x) \cdot f(x|y)\, dx$$

we denote **conditional mean** of the random variable X by

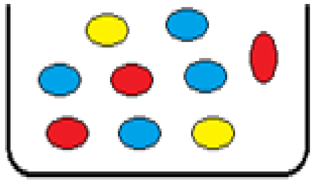
$$\mu_{X|y} = E(X|y)$$

Correspondingly, the **conditional variance** of X given $Y = y$ is

$$\sigma^2_{X|y} = E[(X - \mu_{X|y})^2|y] = E(X^2|y) - \mu^2_{X|y}$$

EXAMPLE

With reference to Example



find the conditional mean of X given $Y = 1$.

		$x = 0$	$x = 1$	$x = 2$	
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{7}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$		$\frac{7}{18}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$	

$$f(x|y) = \frac{f(x, y)}{h(y)} \qquad h(y) \neq 0$$

$$f(0|1) = \frac{\frac{2}{9}}{\frac{7}{18}} = \frac{4}{7}$$

$f(0|1) = \frac{4}{7}$, $f(1|1) = \frac{3}{7}$, and $f(2|1) = 0$

$$E(X|1) = 0 \cdot f(0|1) + 1 \cdot f(1|1) + 2 \cdot f(2|1)$$

$$E(X|1) = 0 \cdot \frac{4}{7} + 1 \cdot \frac{3}{7} + 2 \cdot 0 = \frac{3}{7}$$

THE THEORY IN PRACTICE

The analog of the first moment, $\mu_1' = \mu$, is the **sample mean**, $\bar{x}$, defined as

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

where $i = 1, 2, \ldots, n$ and n is the number of observations.

The **dispersion** of data also is important in its description.

The **sample standard deviation**, s, is

calculated analogously to the second moment, as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x - \overline{x})^2}{n - 1}}$$

Since this formula requires first the calculation of the mean, then subtraction of the mean from each observation before squaring and adding, it is much easier to use the following *calculating formula* for s:

$$s = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n - 1)}}$$

Moment Generating Functions

MGF is an alternative route to analytical results compared with working directly with probability density functions.

The moment-generating function of a random variable X is

$$M_X(t) := \mathrm{E}\left[e^{tX}\right], \quad t \in \mathbb{R},$$

wherever this expectation exists.

The moment-generating function is so named because it can be used to find the moments of the distribution.

The series expansion of e^{tX} is

$$e^{t\,X} = 1 + t\,X + \frac{t^2\,X^2}{2!} + \frac{t^3\,X^3}{3!} + \cdots + \frac{t^n\,X^n}{n!} + \cdots.$$

Hence

$$M_X(t) = \mathrm{E}(e^{t\,X}) = 1 + t\,\mathrm{E}(X) + \frac{t^2\,\mathrm{E}(X^2)}{2!} + \frac{t^3\,\mathrm{E}(X^3)}{3!} + \cdots + \frac{t^n\,\mathrm{E}(X^n)}{n!} + \cdots$$

$$= 1 + tm_1 + \frac{t^2 m_2}{2!} + \frac{t^3 m_3}{3!} + \cdots + \frac{t^n m_n}{n!} + \cdots,$$

Differentiating $M_X(t)$ i times with respect to t and setting $t = 0$, we obtain the ith moment about the origin, m_i;

where m_n is the nth moment.

$$E[X^k] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}.$$

SPECIAL PROBABILITY DISTRIBUTIONS

THE DISCRETE UNIFORM DISTRIBUTION

If a random variable can take on k different values with equal probability, we say that it has a **discrete uniform distribution**;

Definition A random variable X has a **discrete uniform distribution** and it is referred to as a discrete uniform random variable if and only if its probability distribution is given by

$$f(x) = \frac{1}{k} \qquad \text{for } x = x_1, x_2, \ldots, x_k$$

where $x_i \neq x_j$ when $i \neq j$.

The mean and the variance of this distribution are

$$\mu = \sum_{i=1}^{k} x_i \cdot \frac{1}{k} \text{ and } \sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 \cdot \frac{1}{k}.$$

THE BERNOULLI DISTRIBUTION

If an experiment has two possible outcomes, "success" and "failure," and their probabilities are, respectively, θ and $1-\theta$, then the number of successes, 0 or 1, has a **Bernoulli distribution**;

Definition A random variable X has a **Bernoulli distribution** and it is referred to as a Bernoulli random variable if and only if its probability distribution is given by

$$f(x;\theta) = \theta^x(1-\theta)^{1-x} \qquad \text{for } x = 0, 1$$

Thus, $f(0;\theta) = 1-\theta$ and
$f(1;\theta) = \theta$ are combined into a single formula.

THE BINOMIAL DISTRIBUTION

As we shall see, several random variables arise in connection with repeated trials.

To derive a formula for the probability of getting "x successes in n trials" under the stated conditions, observe that the probability of getting x successes and $n - x$ failures *in a specific order* is $\theta^x(1-\theta)^{n-x}$.

Clearly, the number of ways in which we can select the x trials on which there is to be a success is $\binom{n}{x}$,

and it follows that the desired probability for "x successes in n trials"

is $\binom{n}{x}\theta^x(1-\theta)^{n-x}$.

DEFINITION A random variable X has a **binomial distribution** and it is referred to as a binomial random variable if and only if its probability distribution is given by

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad \text{for } x = 0, 1, 2, \ldots, n$$

EXAMPLE

Find the probability of getting five heads and seven tails in 12 flips of a balanced coin.

Solution Substituting $x = 5$, $n = 12$, and $\theta = \frac{1}{2}$ into the formula for the binomial distribution, we get

$$b\left(5; 12, \frac{1}{2}\right) = \binom{12}{5}\left(\frac{1}{2}\right)^5\left(1-\frac{1}{2}\right)^{12-5} = 0.193$$

EXAMPLE:

Show that the mean of the Binomial distribution is $\mu = n\theta$.

$$\mu = \sum_{x=0}^{n} x \cdot \binom{n}{x} \theta^x (1-\theta)^{n-x} = \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} \theta^x (1-\theta)^{n-x}$$

$$= n\theta \cdot \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} \theta^{x-1} (1-\theta)^{n-x}$$

$$= n\theta \cdot \sum_{x=1}^{n} \binom{n-1}{x-1} \theta^{x-1} (1-\theta)^{n-x}$$

letting $y = x - 1$ and $m = n - 1$, this becomes

$$\mu = n\theta \cdot \sum_{y=0}^{m} \binom{m}{y} \theta^y (1-\theta)^{m-y} = n\theta$$

since the last summation is the sum of all the values of a binomial distribution with the parameters m and θ, and hence equal to 1.

EXERCISE:

Show that the variance of the Binomial distribution is $\sigma^2 = n\theta(1-\theta)$.

EXAMPLE:

We toss a coin four times, and that the coin is fair.
Find the probability distribution.

we have $n = 4$ and the probability of tossing a head, say, is $p = \frac{1}{2}$.

If X represents the number of heads in the four tosses,
$X = 0, 1, 2, 3, 4.$

$$P(X=0) = \binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = 1 \times 1 \times \frac{1}{16} = \frac{1}{16},$$

$$P(X=1) = \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = 4 \times \frac{1}{2} \times \frac{1}{8} = \frac{4}{16},$$

$$P(X=2)=\binom{4}{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{2}=6\times\frac{1}{4}\times\frac{1}{4}=\frac{6}{16},$$

$$P(X=3)=\binom{4}{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{1}=4\times\frac{1}{8}\times\frac{1}{2}=\frac{4}{16},$$

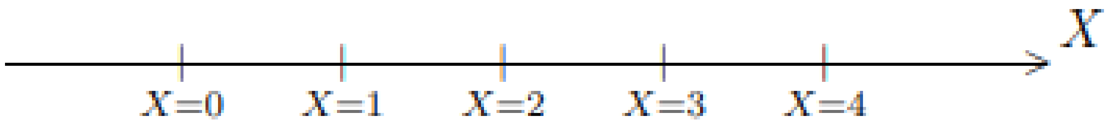
$$P(X=4)=\binom{4}{4}\left(\frac{1}{2}\right)^{4}\left(\frac{1}{2}\right)^{0}=1\times\frac{1}{16}\times 1=\frac{1}{16}.$$

We should note here that the sum of all possible probabilities is one.

$$P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)$$

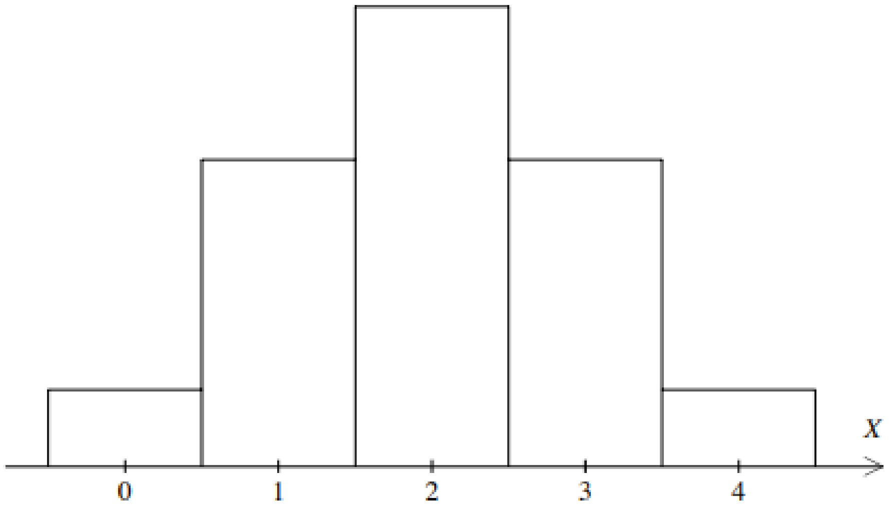
$$\begin{aligned}
&= \frac{1}{16}+\frac{4}{16}+\frac{6}{16}+\frac{4}{16}+\frac{1}{16}\\
&= \frac{16}{16}\\
&= 1.
\end{aligned}$$

It is very useful to represent these probabilities in a diagram. We will represent the possible values that the random variable X can take along a horizontal axis:



$$P(X=0)=\frac{1}{16}, \qquad P(X=1)=\frac{4}{16}, \qquad P(X=2)=\frac{6}{16},$$

$$P(X=3)=\frac{4}{16}, \qquad \text{and} \qquad P(X=4)=\frac{1}{16}.$$



The block above $X = 0$ is represented as one unit in width from $-\frac{1}{2}$ to $\frac{1}{2}$, while the block above $X = 1$ will be one unit wide from $\frac{1}{2}$ to $1\frac{1}{2}$.

In this case, each block will be the same width of one unit.

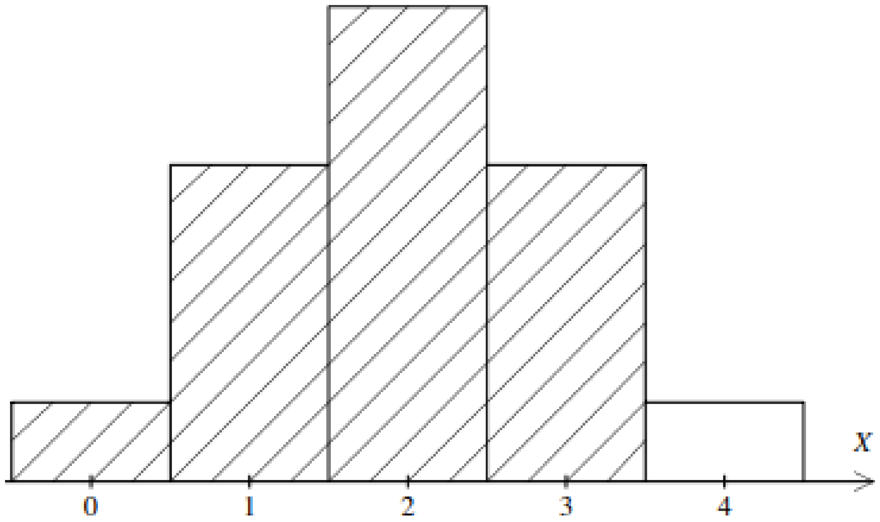
Now the area covered by each block will represent the probability for the corresponding value of X.

If we wanted to calculate the probability of obtaining three heads or less, then we would have to calculate the four probabilities for $X = 0, 1, 2, 3$

The probability of obtaining three heads or less in four tosses of a fair coin

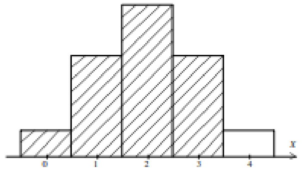
$$
\begin{aligned}
P(X \leq 3) &= P(X=0) + P(X=1) + P(X=2) \\
&\quad + P(X=3) \\
&= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} \\
&= \frac{15}{16} \\
&= 0.9375.
\end{aligned}
$$

This probability can also be represented by the shaded area given below:

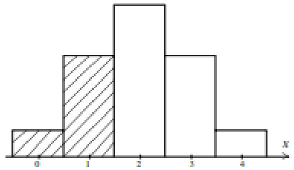


We can see this simply from diagrams:

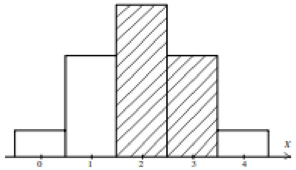
$P(X \leq 3)$



$P(X \leq 1)$



$P(2 \leq X \leq 3)$



THE POISSON DISTRIBUTION

When n is large, the calculation of binomial probabilities with the formula will usually involve a prohibitive amount of work.

When $n \to \infty, \theta \to 0$, while $n\theta$ remains constant we can write $n\theta = \lambda$.

Letting this constant be λ, hence, $\theta = \frac{\lambda}{n}$,

$$b(x; n, \theta) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

After some algebra and approximations;

the distribution becomes

$$p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \text{for } x = 0, 1, 2, \ldots$$

Definition A random variable X has a **Poisson distribution** and it is referred to as a Poisson random variable if and only if its probability distribution is given by

$$p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad \text{for } x = 0, 1, 2, \ldots$$

Thus, in the limit when $n \to \infty$, $\theta \to 0$, and $n\theta = \lambda$ remains constant, the number of successes is a random variable having a Poisson distribution with the parameter λ.

EXAMPLE:

Births in a hospital occur randomly at an average rate of 1.8 births per hour.

What is the probability of observing 4 births in a given hour at the hospital?

Let $X =$ # of births in a given hour

(i) Events occur randomly
(ii) Mean rate $\lambda = 1.8$

$\Rightarrow X \sim \mathrm{Po}(1.8)$

We can now use the formula to calculate the probability of observing exactly 4 births in a given hour

$$P(X = 4) = \mathrm{e}^{-1.8} \frac{1.8^4}{4!} = 0.0723$$

What about the probability of observing more than or equal to 2 births in a given hour at the hospital?

We want $P(X \geq 2) = P(X = 2) + P(X = 3) + \ldots$

i.e. an infinite number of probabilities to calculate but

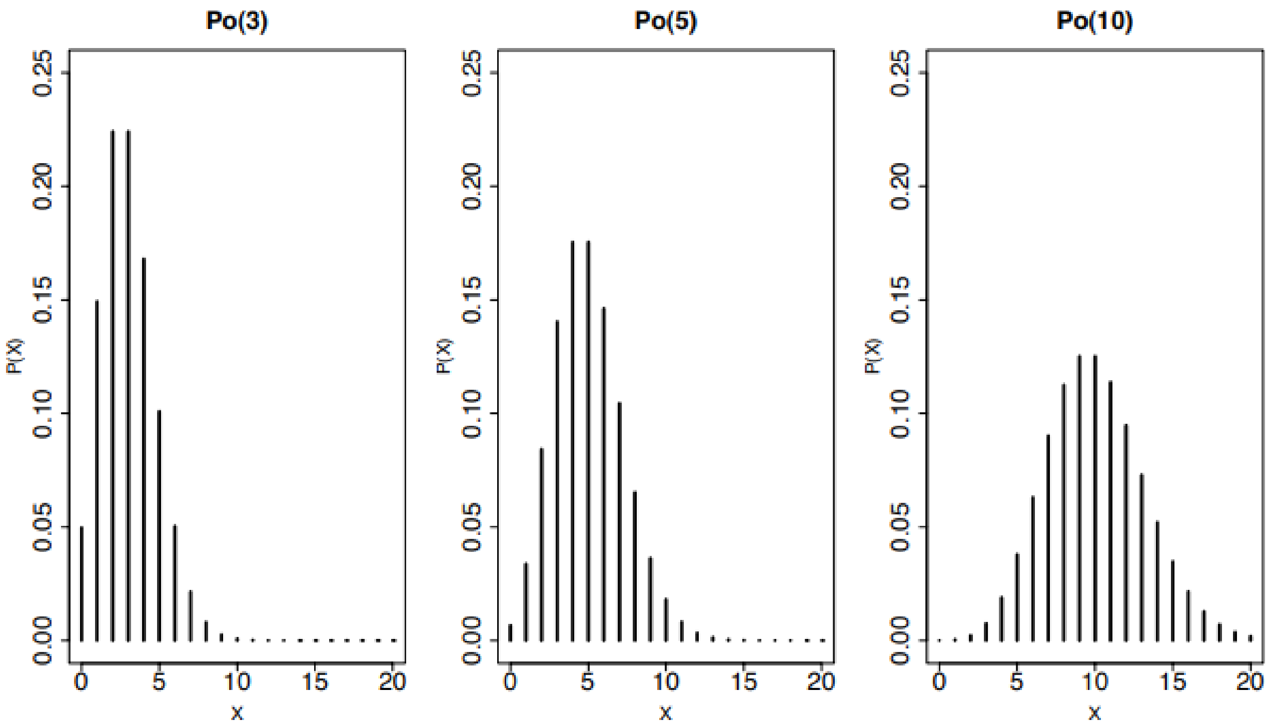
$$\begin{aligned} P(X \geq 2) &= P(X = 2) + P(X = 3) + \ldots \\ &= 1 - P(X < 2) \end{aligned}$$

$$= 1 - (P(X = 0) + P(X = 1))$$

$$= 1 - \left(e^{-1.8}\frac{1.8^0}{0!} + e^{-1.8}\frac{1.8^1}{1!} \right)$$

$$\begin{aligned} &= 1 - (0.16529 + 0.29753) \\ &= 0.537 \end{aligned}$$

The shape of the Poisson distribution



We observe that the Poisson distributions

decreases as λ increases are centred roughly on λ;

have variance (spread) that increases as λ increases.

THEOREM The mean and the variance of the Poisson distribution are given by

$$\mu = \lambda \quad \text{and} \quad \sigma^2 = \lambda$$

THEOREM The moment-generating function of the Poisson distribution is given by

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Proof. By Definition

$$M_X(t) = \sum_{x=0}^{\infty} e^{xt} \cdot \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \cdot \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

where $\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$ can be recognized as the series

of e^z with $z = \lambda e^t$.

Thus,

$$M_X(t) = e^{-\lambda} \cdot e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Then, if we differentiate $M_X(t)$ twice with respect to t, we get

$$M'_X(t) = \lambda e^t e^{\lambda(e^t - 1)}$$
$$M''_X(t) = \lambda e^t e^{\lambda(e^t - 1)} + \lambda^2 e^{2t} e^{\lambda(e^t - 1)}$$

$$\mu'_1 = M'_X(0) = \lambda \qquad \text{and } \mu'_2 = M''_X(0) = \lambda + \lambda^2.$$

Thus, $\mu = \lambda$ and $\sigma^2 = \mu'_2 - \mu^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda$.