Week 9

Oct. 29 – Nov. 2

Important distributions – discrete and continuous

Lecture 23.

Negative Binomial and Geometric.

Let T_1, T_2,..., T_n be independent and identically distributed (i. i. d.) with

$$T_i \sim Geometric\,(p)$$

i.e.,

$$P\,(T_1 = x) = q^{x-1}p,\ x = 1, 2, 3, \ldots$$

Then the distribution of $T = T_1 + T_2 + \ldots + T_n$ is $Negative\ Binomial\,(n, p)$ with

$$\begin{aligned} P\,(T = x) &= \binom{x-1}{n-1} p^n q^{x-n},\ x \geq n \\ ET &= n/p \\ Var\,(T) &= nq/p^2. \end{aligned}$$

Poisson Distribution.

$Poisson\,(\lambda)$

$$P\,(X = k) = \frac{\exp\,(-\lambda)\,\lambda^k}{k!},\ x = 0, 1, 2, 3, \ldots$$

Fact:

$$\begin{aligned} e^\lambda &= 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots + \frac{\lambda^k}{k!} + \cdots \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \end{aligned}$$

This implies

$$\sum_{k=0}^{\infty} \exp\,(-\lambda)\,\frac{\lambda^k}{k!} = ?$$

Typical Poisson variables are the number of raisins in a cake, flies in a room, fumbles in a football game, car accidents on a street corner, murders in a town, number of bomb hits in South London during World War II.

Question:

$$\begin{aligned} EX &= \lambda? \\ E\,(X - EX)^2 &= \lambda? \end{aligned}$$

Solution:

Expectation

$$
\begin{aligned}
EX &= \sum_{k=0}^{\infty} k \frac{\exp(-\lambda)\,\lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{\exp(-\lambda)\,\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} \frac{\exp(-\lambda)\,\lambda^k}{(k-1)!} = \lambda \sum_{k=1}^{\infty} \frac{\exp(-\lambda)\,\lambda^{k-1}}{(k-1)!} \\
&= \lambda \sum_{j=0}^{\infty} \frac{\exp(-\lambda)\,\lambda^j}{j!} =?
\end{aligned}
$$

Variance

$$
\begin{aligned}
Var(X) &= EX^2 - (EX)^2 \\
&= \sum_{k=0}^{\infty} k^2 \frac{\exp(-\lambda)\,\lambda^k}{k!} - \lambda^2 \\
&= \sum_{k=1}^{\infty} k \frac{\exp(-\lambda)\,\lambda^k}{(k-1)!} - \lambda^2 \\
&= \sum_{j=0}^{\infty} (j+1) \frac{\exp(-\lambda)\,\lambda^{j+1}}{j!} - \lambda^2 \\
&= \lambda \sum_{j=0}^{\infty} j \frac{\exp(-\lambda)\,\lambda^j}{j!} + \lambda - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned}
$$

Example: As an example of a Poisson random variable consider the statistics of flying bomb hits in the south of London during World War II. The entire area is divided into a grid of $N = 576$ small areas of size one quarter square kilometer each. The table below records the number of squares with $0, 1, 2, 3$ etc. hits each. The total number of hits is 537. The average number of hits per square is then $537/576 = .93$ hits per square. It can be shown that if the targeting is completely random, then the probability that a square is hit with 0,1,2,3 etc. hits is governed by a Poisson distribution.

Number of hits k	0	1	2	3	4	5 or more
Number of areas, k hits	229	211	93	35	7	1
$576 \cdot P(X = k)$, $X \sim Po(.93)$	227	211	99	31	7	2

Question: Let X_1, X_2,..., X_n be independent and identically distributed (i. i. d.) with

$$X_i \sim Poisson(\lambda),$$

i.e.,

$$P(X = k) = \frac{\exp(-\lambda)\,\lambda^k}{k!},\ x = 0, 1, 2, 3, \ldots$$

Let $X = X_1 + X_2 + \ldots + X_n$. Then

$$\begin{aligned} EX &= ?, Var(X) = ? \\ E\left(\frac{X - n\lambda}{\sqrt{n\lambda}}\right) &= ?, Var\left(\frac{X - n\lambda}{\sqrt{n\lambda}}\right) = ? \end{aligned}$$

and What is the distribution of $X = X_1 + X_2 + \ldots + X_n$?

Solution: (i)

$$\begin{aligned} EX &= n\lambda, Var(X) = n\lambda \\ E\left(\frac{X - n\lambda}{\sqrt{n\lambda}}\right) &= 0, Var\left(\frac{X - n\lambda}{\sqrt{n\lambda}}\right) = 1 \end{aligned}$$

(ii) Result for sum of Poisson variables,

$$X_1 \sim Poisson(\lambda_1), \, X_2 \sim Poisson(\lambda_2) \Longrightarrow X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2)$$

Why?

$$\begin{aligned} P(X_1 + X_2 = k) &= \sum_{j=0}^{k} P(X_1 = j) P(X_2 = k - j) \\ &= \sum_{j=0}^{k} \frac{\exp(-\lambda_1) \lambda_1^j}{j!} \frac{\exp(-\lambda_2) \lambda_2^{k-j}}{(k-j)!} \\ &= \frac{\exp(-\lambda_1 - \lambda_2)}{k!} \sum_{j=0}^{k} k! \frac{\lambda_1^j}{j!} \frac{\lambda_2^{k-j}}{(k-j)!} \\ &= \frac{\exp(-\lambda_1 - \lambda_2)}{k!} \sum_{j=0}^{k} \binom{n}{k} \lambda_1^j \lambda_2^{k-j} = ? \end{aligned}$$

Fact: $(a+b)^k = (a+b) \cdot (a+b) \cdot \cdots \cdot (a+b) = \sum_{j=0}^{k} \binom{n}{k} a^j b^{k-j}$

So

$$X = X_1 + X_2 + \ldots + X_n = Po(n\lambda).$$

Connections to Binomial distribution.

Question:

$$P(X_1 = j | X_1 + X_2 = k) = \frac{\frac{\exp(-\lambda_1)\lambda_1^j}{j!} \frac{\exp(-\lambda_2)\lambda_2^{k-j}}{(k-j)!}}{\frac{\exp(-\lambda_1 - \lambda_2)}{k!} (\lambda_1 + \lambda_2)^k} = \binom{k}{j} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^j \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{k-j}$$

i.e.,

$$X_1 \sim Binomial\left(k, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

Connection 2: $Binomial(n, p)$

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, \ldots, n.$$

Assume that

$$\lim_{n\to\infty} np = \lambda,$$

then

$$\begin{aligned}
& \lim_{n\to\infty} P(X=k) \\
= & \lim_{n\to\infty} \frac{n(n-1)\cdot\cdots\cdot(n-k+1)}{k!} p^k (1-p)^{n-k} \\
= & \lim_{n\to\infty} \frac{np\cdot(n-1)p\cdot\cdots\cdot(n-k+1)p}{k!} \exp\left[-(n-k)p\cdot\frac{\log(1-p)}{-p}\right] \\
= & \frac{\exp(-\lambda)\lambda^k}{k!}
\end{aligned}$$

Connection to Binomial distribution: Assume that

$$\lim_{N\to\infty} \frac{k}{N} = p,$$

then

$$\lim_{N\to\infty} P(X=x) = \binom{n}{x} p^x q^{n-x}?$$

Multinomial – An extension of Binomial

$$\begin{aligned}
(X_1, X_2, \ldots, X_k) &\sim Multinomial(n, p_1, p_2, \ldots, p_k), \\
p_1 + p_2 + \ldots + p_k &= 1,\ p_i \ge 0, \\
P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) &= \frac{n!}{x_1! x_2! \cdot \ldots \cdot x_k!} p_1^{x_1} p_2^{x_2} \cdot \ldots \cdot p_k^{x_k}, \\
x_1 + x_2 + \ldots + x_k &= n,\ x_i \ge 0,
\end{aligned}$$

Suppose an experiment is performed n times, and the result on each occasion is one of k types. It is of type i with probability p_i. Let X_i be the number of results of type i in n trials. Then $(X_1, X_2, \ldots, X_k) \sim Multinomial(n, p_1, p_2, \ldots, p_k)$. For example, toss a die n times; then $(X_1, X_2, \ldots, X_6) \sim Multinomial\left(n, \frac{1}{6}, \frac{1}{6}, \ldots, \frac{1}{6}\right)$ is the number of times $1, 2, 3, 4, 5, 6$ appeared in n times.

Multinomial and its connection to Poisson

Similar to $k = 2$, we have the following result.

Let $X_1 \sim Poisson(\lambda_1)$, $X_2 \sim Poisson(\lambda_2)$, $\ldots$, $X_k \sim Poisson(\lambda_k)$ independent. Then

$$(X_1, X_2, \ldots, X_k) \,|\, X_1 + X_2 + \ldots + X_k = n \sim Multinomial(n, p_1, p_2, \ldots, p_k)$$

with

$$p_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \ldots + \lambda_n}$$

Hypergeometric.

Suppose that we have a set of N balls, of which k are red and $N-k$ are blue. We choose n balls, without replacement, and define X to be the number of the red balls in our sample.

$$P(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

Connection to Binomial distribution: Assume that

$$\lim_{N\to\infty}\frac{k}{N}=p,\ 0<p<1$$

then

$$\lim_{N\to\infty}P(X=x)=\binom{n}{x}p^{x}q^{n-x}?$$

Example 5.6 in the textbook.
Observed tables

	Democrat(*yellow*)	Republican(*red*)	
Female	24	4	28
Male	8	14	22
	32	18	50

Expected table

	Democrat (*yellow*)	Republican(*red*)	
Female	18	10	28
Male	14	8	22
	32	18	50

If we choose 28 balls out of 50, we should expect to see (under the assumption the traits of gender and political party are independent), on the average, the same percentage of yellow balls in our sample as in the urn, if there is no association between gender and political party. Thus we see expect to see $28\frac{32}{50}\approx 17.92\approx 18$ yellow balls in our sample. But the actual number in our data is 24. What is the probability to have an observation ≥ 24? It is just .000395.

Lecture 24. Normal distribution
The most important density function

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

There is a special notation for the density $\phi(x)$ and for the cumulative distribution function $\Phi(x)$.

It is even not easy to see why

$$\int \phi(x)\,dx = 1?$$

Trick:

$$\begin{aligned}\left(\int \phi(x)\,dx\right)^2 &= \int \phi(x)\,dx \int \phi(y)\,dy \\ &= \frac{1}{2\pi}\int\int \exp\left(-\frac{1}{2}\left(x^2+y^2\right)\right) dxdy\end{aligned}$$

Now switch to polar coordinates

$$x = r\cos\theta,\ y = r\sin\theta,$$

$$\left(\int \phi(x)\,dx\right)^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 \exp\left(-\frac{1}{2}r^2\right) rdrd\theta = \frac{1}{2\pi}\int_0^{2\pi}\left[-\exp\left(-\frac{1}{2}r^2\right)\right]_0^\infty d\theta = 1.$$

Question:

$$EX?$$

Question:

$$\begin{aligned}EX^2 &= ? \\ Var(X) &= ?\end{aligned}$$

Question: Let $Y = \mu + X\sigma$ with $X \sim N(0,1)$. Then

$$\begin{aligned}EY &= ? \\ Var(Y) &= ? \\ f_Y(x) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\end{aligned}$$

More generally, we say $Y \sim N\left(\mu,\sigma^2\right)$ if $\frac{Y-\mu}{\sigma} \sim N(0,1)$.

Normal distribution – An approximation of Binomial by Abraham de Moivre

Let $X \sim Binomial(n,p)$. It can be shown that

$$P\left(a \le \frac{X-np}{\sqrt{npq}} \le b\right) \approx \int_a^b \phi(x)\,dx$$

Lecture 25. Sum of independent Normal random variables

Let $X \sim N\left(0, \sigma_X^2\right)$ and $Y \sim N\left(0, \sigma_Y^2\right)$ be independent. The joint density of (X, Y) is

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{x^2}{2\sigma_X^2} - \frac{y^2}{2\sigma_Y^2}\right).$$

What is joint density of (Y, Z) with $Z = X + Y$?

$$\begin{aligned} f_{X,Z}(x, z) &= f_{Z|Y}(z|y) f_Y(y) \\ &= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{y^2}{2\sigma_Y^2} - \frac{(z-y)^2}{2\sigma_X^2}\right) \end{aligned}$$

Theorem. More generally, let f_X and f_Y be the densities of two independent random variables X and Y respectively. Let $Z = X + Y$, then

$$\begin{aligned} f_{X,Z}(x, z) &= f_{Z|X}(z|y) f_Y(y) \\ &= f_X(z-y) f_Y(y) \end{aligned}$$

and the marginal density of Z is

$$f_Z(z) = \int f_X(z-y) f_Y(y)\, dy$$

Definition. The convolution $f * g$ of f and g is the function given by

$$f * g = \int f(z-y)\, g(y)\, dy.$$

Let $X \sim N\left(0, \sigma_X^2\right)$ and $Y \sim N\left(0, \sigma_Y^2\right)$ be independent. The marginal density of Z is

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi\sigma_X\sigma_Y} \int \exp\left(-\frac{y^2}{2\sigma_Y^2} - \frac{(z-y)^2}{2\sigma_X^2}\right) dy \\ &= \frac{1}{2\pi\sigma_X\sigma_Y} \int \exp\left(-\frac{y^2}{2}\left(\frac{1}{\sigma_Y^2} + \frac{1}{\sigma_X^2}\right) + \frac{zy}{\sigma_X^2} - \frac{z^2}{2\sigma_X^2}\right) dy \end{aligned}$$

Note that

$$\int \exp\left[-a\frac{(y-b)^2}{2} + c\right] dy = \frac{\sqrt{2\pi}}{\sqrt{a}} e^c$$

i.e.,

$$\int \exp\left[-\frac{a}{2}y^2 + yab - \frac{ab^2}{2} + c\right] dy = \frac{2\pi}{\sqrt{a}} e^c.$$

Let

$$
\begin{aligned}
a &= \frac{1}{\sigma_Y^2} + \frac{1}{\sigma_X^2} = \frac{\sigma_X^2 + \sigma_Y^2}{\sigma_X^2 \sigma_Y^2} \\
b &= \frac{z\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2} \\
c &= -\frac{z^2}{2\sigma_X^2} + \frac{1}{2}\frac{\sigma_X^2 + \sigma_Y^2}{\sigma_X^2 \sigma_Y^2} \cdot \left(\frac{z\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}\right)^2 \\
&= -\frac{z^2}{2\sigma_X^2} + \frac{1}{2\sigma_X^2} \cdot \frac{z^2\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2} = -\frac{1}{2}\frac{z^2}{\sigma_X^2 + \sigma_Y^2}
\end{aligned}
$$

then

$$
\begin{aligned}
f_Z(z) &= \frac{\sigma_X \sigma_Y}{\sqrt{2\pi}\sigma_X \sigma_Y \cdot \sqrt{\sigma_X^2 + \sigma_Y^2}} \exp\left(-\frac{1}{2}\frac{z^2}{\sigma_X^2 + \sigma_Y^2}\right) \\
&= \frac{\sigma_X \sigma_Y}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2}} \exp\left(-\frac{1}{2}\frac{z^2}{\sigma_X^2 + \sigma_Y^2}\right).
\end{aligned}
$$

Thus

$$
X \sim N\left(0, \sigma_X^2\right), Y \sim N\left(0, \sigma_Y^2\right) \text{ independent } \implies X + Y \sim N\left(0, \sigma_X^2 + \sigma_Y^2\right)
$$