Lecture 11

Chi square distribution
Two random variables
Joint & Marginal Probability Functions

Index

- Chi square distribution
- Skewed Distribution
- Chi square curve
- Two discrete random variables
- Marginal probability mass function
- Two continuous random variables
- Marginal probability density function
- Independent random variables
- Questions

Chi square Distribution

- The chi-squared distribution is important because it is the basis for a number of procedures in statistical inference.
- Gamma pdf is

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & x\geq 0\\ 0 & \text{otherwise}\end{cases}$$

Chi square Distribution

Let ν be a positive integer. Then a random variable X is said to have a **chi-squared distribution** with parameter ν if the pdf of X is the gamma density with $\alpha = \nu/2$ and $\beta = 2$. The pdf of a chi-squared rv is thus

$$f(x;\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

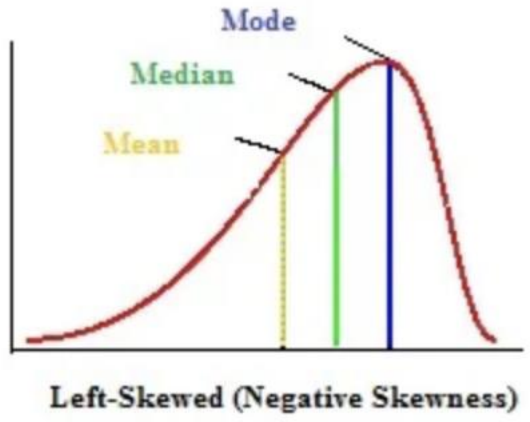
The parameter ν is called the **number of degrees of freedom** (df) of X. The symbol χ^2 is often used in place of "chi-squared."

Skewed Distribution

- If one tail is longer than another, the distribution is skewed.
- These distributions are sometimes called asymmetric or asymmetrical distributions as they don’t show any kind of symmetry.
- In a normal distribution, the mean and the median are the same number while the mean and median in a skewed distribution become different numbers.

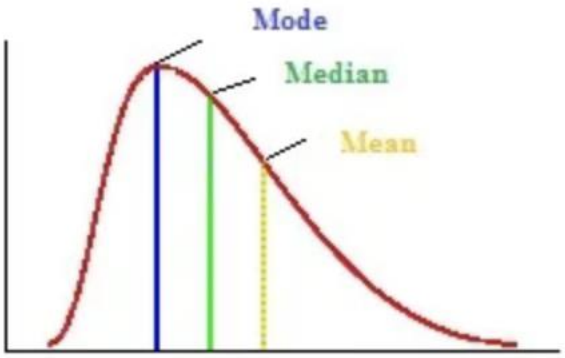
Left skewed distribution

- A left-skewed distribution has a long left tail.
- Left-skewed distributions are also called negatively-skewed distributions.
- That’s because there is a long tail in the negative direction on the number line.
- Mean is also to the left of the peak.
- Mean is to the left of the median also.



Right skewed Distribution

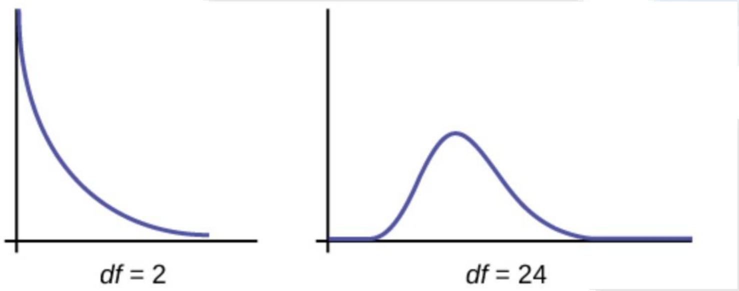
- A right-skewed distribution has a long right tail.
- Right-skewed distributions are also called positive-skew distributions.
- That’s because there is a long tail in the positive direction on the number line.
- Mean is to the right of the peak.
- Mean is to the right of the median also.



Right-Skewed (Positive Skewness)

Chi square curve

- The curve is nonsymmetrical and skewed to the right.
- There is a different chi-square curve for each dof.
- The mean is located to the right of the peak.
- Mean = dof
- Variance = 2*dof



Two discrete random variables

Let X and Y be two discrete rv's defined on the sample space $\mathcal{S}$ of an experiment. The **joint probability mass function** $p(x, y)$ is defined for each pair of numbers (x, y) by

$$p(x, y) = P(X = x \text{ and } Y = y)$$

It must be the case that $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$.

Marginal Probability Mass Function

The **marginal probability mass function of *X*,** denoted by $p_X(x)$, is given by

$$p_X(x) = \sum_{y:\, p(x,y)>0} p(x, y) \quad \text{for each possible value } x$$

Similarly, the **marginal probability mass function of *Y*** is

$$p_Y(y) = \sum_{x:\, p(x,y)>0} p(x, y) \quad \text{for each possible value } y.$$

Question

A large insurance agency services a number of customers who have purchased both a homeowner's policy and an automobile policy from the agency. For each type of policy, a deductible amount must be specified. For an automobile policy, the choices are $100 and $250, whereas for a homeowner's policy, the choices are 0, $100, and $200. Suppose an individual with both types of policy is selected at random from the agency's files. Let X be the deductible amount on the auto policy and Y be the deductible amount on the homeowner's policy. Suppose the joint pmf is given as follows:

		y		
$p(x, y)$		0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

Question

Calculate the following:

1. p(100,100)
2. P(Y>=100)
3. $p_X(x)$
4. $p_Y(y)$

$p(x, y)$		y: 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

1. p(100,100) = P(X=100 and Y =100) = 0.1
2. P(Y>=100) = p(100,100) + p(250,100) + p(100,250) + p(250,200) = 0.75 (Can be computed from pmf of Y too)

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

$$p_X(100) = p(100, 0) + p(100, 100) + p(100, 200) = .50$$

and

$$p_X(250) = p(250, 0) + p(250, 100) + p(250, 200) = .50$$

The marginal pmf of X is then

$$p_X(x) = \begin{cases} .5 & x = 100, 250 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, the marginal pmf of Y is obtained from column totals as

$$p_Y(y) = \begin{cases} .25 & y = 0, 100 \\ .50 & y = 200 \\ 0 & \text{otherwise} \end{cases}$$

Question

A service station has both self-service and full-service islands. On each island, there is a single regular unleaded pump with two hoses. Let X denote the number of hoses being used on the self-service island at a particular time, and let Y denote the number of hoses on the full-service island in use at that time. The joint pmf of X and Y appears in the accompanying tabulation.

Question

$p(x, y)$		y: 0	1	2
	0	.10	.04	.02
x	1	.08	.20	.06
	2	.06	.14	.30

a. What is $P(X = 1 \text{ and } Y = 1)$?

b. Compute $P(X \leq 1 \text{ and } Y \leq 1)$.

c. Give a word description of the event $\{X \neq 0 \text{ and } Y \neq 0\}$, and compute the probability of this event.

d. Compute the marginal pmf of X and of Y. Using $p_X(x)$, what is $P(X \leq 1)$?

Answers

a. 0.20
b. 0.42
c. Atleast one of the hoses is there in both full service and self service islands, 0.7
d. $p_x(0) = 0.16$, $p_x(1) = 0.34$, $p_x(2) = 0.5$, 0 otherwise
   $p_y(0) = 0.24$, $p_1(1) = 0.38$, $p_2(0) = 0.38$, 0 otherwise
   $P(X<=1) = 0.5$

Two continuous random variables

Let X and Y be continuous rv's. A **joint probability density function** $f(x, y)$ for these two variables is a function satisfying $f(x, y) \geq 0$ and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$. Then for any two-dimensional set A

$$P[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy$$

Marginal Probability Density Function

The **marginal probability density functions** of X and Y, denoted by $f_X(x)$ and $f_Y(y)$, respectively, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad \text{for } -\infty < y < \infty$$

Independent random variables

Two random variables X and Y are said to be **independent** if for every pair of x and y values

$$p(x, y) = p_X(x) \cdot p_Y(y) \qquad \text{when } X \text{ and } Y \text{ are discrete}$$

or

$$f(x, y) = f_X(x) \cdot f_Y(y) \qquad \text{when } X \text{ and } Y \text{ are continuous}$$

For the given pdf

$$f(x,y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Verify it is a legitimate pdf.
2. Find out $P(0 \le X \le 0.25, 0 \le Y \le 0.25)$
3. Find marginal pdf of X
4. Find marginal pdf of Y
5. Find $P(0.25 \le Y \le 0.75)$

Answer 1.

To verify that this is a legitimate pdf, note that $f(x, y) \geq 0$ and

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy &= \int_0^1\int_0^1 \frac{6}{5}(x + y^2)\,dx\,dy \\ &= \int_0^1\int_0^1 \frac{6}{5}x\,dx\,dy + \int_0^1\int_0^1 \frac{6}{5}y^2\,dx\,dy \\ &= \int_0^1 \frac{6}{5}x\,dx + \int_0^1 \frac{6}{5}y^2\,dy = \frac{6}{10} + \frac{6}{15} = 1\end{aligned}$$

Answer 2.

$$
\begin{aligned}
P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) &= \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x + y^2)\,dx\,dy \\
&= \frac{6}{5}\int_0^{1/4}\int_0^{1/4} x\,dx\,dy + \frac{6}{5}\int_0^{1/4}\int_0^{1/4} y^2\,dx\,dy \\
&= \frac{6}{20}\cdot \left.\frac{x^2}{2}\right|_{x=0}^{x=1/4} + \frac{6}{20}\cdot \left.\frac{y^3}{3}\right|_{y=0}^{y=1/4} = \frac{7}{640} \\
&= .0109
\end{aligned}
$$

Answer 3 & 4

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 \frac{6}{5}(x + y^2)\, dy = \frac{6}{5}x + \frac{2}{5}$$

for $0 \le x \le 1$ and 0 otherwise. The marginal pdf of Y is

$$f_Y(y) = \begin{cases} \dfrac{6}{5}y^2 + \dfrac{3}{5} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Answer 5

$$P\left(\frac{1}{4} \le Y \le \frac{3}{4}\right)$$

$$= \int_{\frac{1}{4}}^{\frac{3}{4}} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

$$= \int_{\frac{1}{4}}^{\frac{3}{4}} f_Y(y)\,dy = \int_{\frac{1}{4}}^{\frac{3}{4}} \frac{6y^2}{5} + \frac{3}{5}$$

$$= \frac{1}{5}\int_{\frac{1}{4}}^{\frac{3}{4}} (6y^2+3)\,dy = \frac{1}{5}\left[\frac{6y^3}{3} + 3y\right]_{1/4}^{3/4}$$

$$= \frac{1}{5}\left[2y^3 + 3y\right]_{1/4}^{3/4} = \frac{1}{5}\left[2\cdot\frac{27}{64} + \frac{9}{4} - 2\cdot\frac{1}{64} - \frac{3}{4}\right]$$

$$= \frac{1}{5}\left[\frac{54}{64} + \frac{9}{4} - \frac{2}{64} - \frac{3}{4}\right]$$

$$= \frac{1}{5}\left[\frac{52}{64} + \frac{6}{4}\right] = \frac{1}{5}\left[\frac{52}{64} + \frac{96}{64}\right]$$

$$= \frac{1}{5}\left[\frac{148}{64}\right] = \frac{1}{5}\left[\frac{37}{16}\right] = \frac{37}{80}$$

References

Probability and statistics for engineers RA Johnson, I Miller, JE Freund - 2000 - 117.239.47.98
Statistics for business & economics DR Anderson, DJ Sweeney, TA Williams, JD Camm
Probability and statistics for engineering and science J Deovre