Module 5: Multivariate Normal Distribution

A variable X follows a discrete probability distribution if the possible values of X are either:

- A finite set
- A countable infinite sequence

$p_x(x_i) = P(X=x_i)$ is called the probability mass function (PMF)

- $p_x(x_i)$ >= 0 as it is a probability
- The sum of PMF for all values of X = 1

Recall that in a Discrete Probability Distribution :

Expected value: $E(X) = \sum_{i=1}^{K} x_i p_X(x_i)$

Variance: $V(X) = \sum_{i=1}^{K} \left(x_i - E(X)\right)^2 p_X(x_i)$

Cumulative distribution function (CDF):

$$F_X(x) = P(X \leq x) = \sum_{i:\, x_i \leq x} p_X(x_i)$$

In a Continuous Probability Distribution:

$$f_X(x)dx \approx \mathrm{P}(x < X < x + dx)$$

$$\int_a^b f_X(x)dx = 1$$

$$F_X(x) = \mathrm{P}(X \leq x) = \int_a^x f_X(t)dt \text{ (cumulative dist. function)}$$

$$E(X) = \int_a^b x f_X(x)dx$$

$$V(X) = \int_a^b (x-\mu)^2 f_X(x)dx$$

Because in a discrete set we are not concerned with the values in between our domain values.

Moment Generating Function

Moments are expected values of X, such as E(X), $E(X^2)$ = E(V), $E(X^3)$, etc. This, can also be calculated using the Moment Generating Function (MGF):

$$M_X(t) = E(e^{tX})$$

The rth moment of X, $E(X^r)$ can be obtained by differentiating $M_x(t)$ r times with respect to t and setting t=0

- $M_x(0) = 1$
- $M^I_x(0) = E(X)$
- $M^{II}_x(0) = E(X^2)$ -> $V(X) = M^{II}_x(0) - (MI_x(0))2$
- In general, $M_x^{(r)}(0) = E(X^r)$

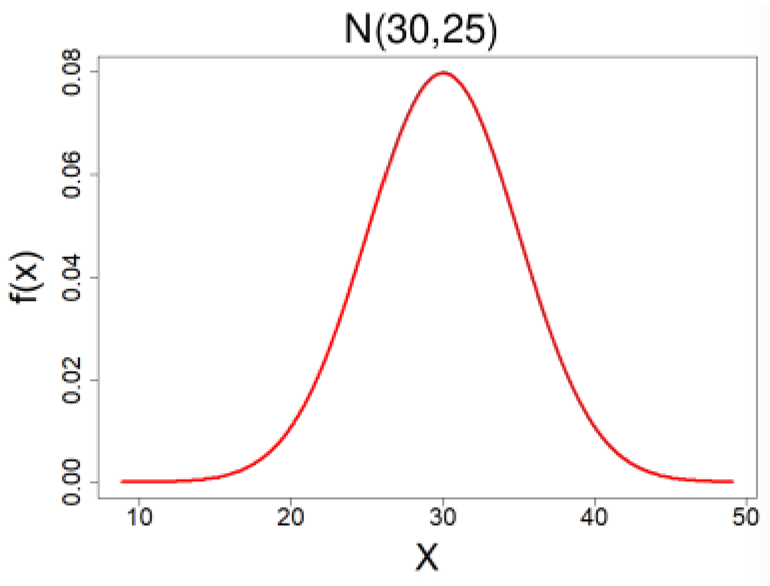
In short, the nth moment is the nth derivative of MGF.

Uniqueness: if X and Y are two random variables and $M_x(t) = M_y(t)$ when |t| < h for some positive number h, then X and Y have the same distribution

Note: MGF does not exist for all distributions ($E(e^{tx})$ may be infinity)

Important Distributions

Normal Distribution



X ~ N(μ, σ^2)   -infinity < μ < infinity , σ < 0

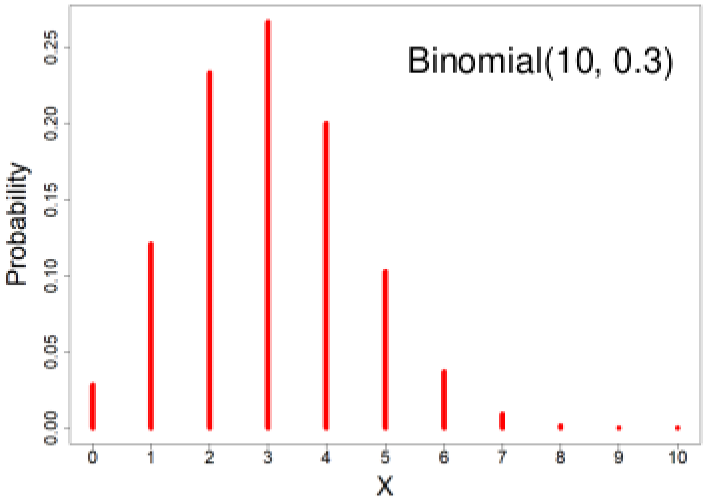
- PDF:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \text{ for } -\infty < x < \infty$$

- E(X) = μ
- V(X) = σ^2
- MGF:

$$M_X(t) = exp(\mu t + \frac{\sigma^2 t^2}{2})$$

Binomial Distribution



X ~ Binomial(n, p)   □□ $\in [0, 1]$

X = the number of successes in n trials when the probability of success in each trail is p.

We can think of X as the sum of n independent Bernoulli(p) random variables, with the same p for every X_i

$$X = \sum_{i=1}^{n} X_i \ ; \ \ X_i = 1 \text{ with probability } p \text{, else } 0; \ \ X_i's \text{ are independent}$$

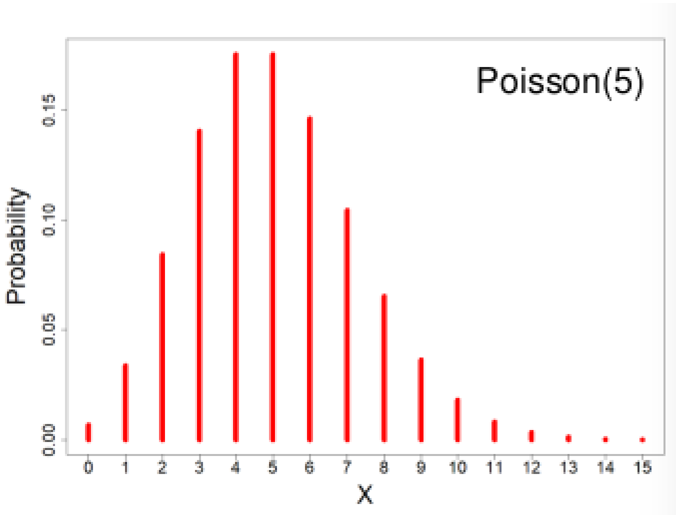
- PMF:

$$\binom{n}{x} p^x (1-p)^{n-x}$$

- Expected value = E(X) = np
- Variance = V(X) = np(1-p)
- MGF = $M_x(t) = (pe^t + (1-p))^n$
- Two discrete random variables are independent if: P(X = x & Y = y) = P(X = x)*P(Y=y)

Ex. A study which analyzed the prevalence of a disease in a population.

Poisson Distribution



X ~ Poisson(λ)   λ > 0

X = The number of occurrences of an event of interest.

- PMF:

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

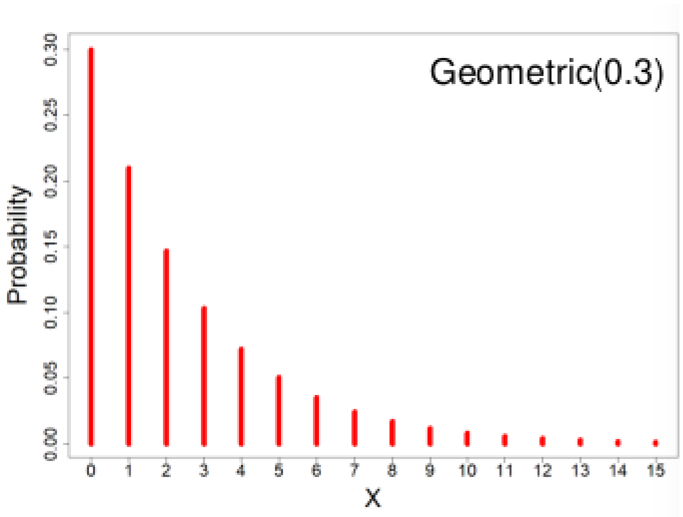
- Expected Values = E(X) = λ
- Variance = V(X) = λ
- MGF = $M_x(t) = e^{\lambda(e^{\wedge}t - 1)}$

Poisson as an approximation of the Binomial Distribution

- If X ~ Binomial(n, p) and n -> infinity, p-> 0 such that np is a constant => X ~ Poisson(np)
- This assumes each event is independent
- Often used analyzing rare diseases

Ex. Analyzing lung cancer in 1000 smokers and non-smokers. This is binomial but can be estimated as a Poisson distribution.

Geometric Distribution



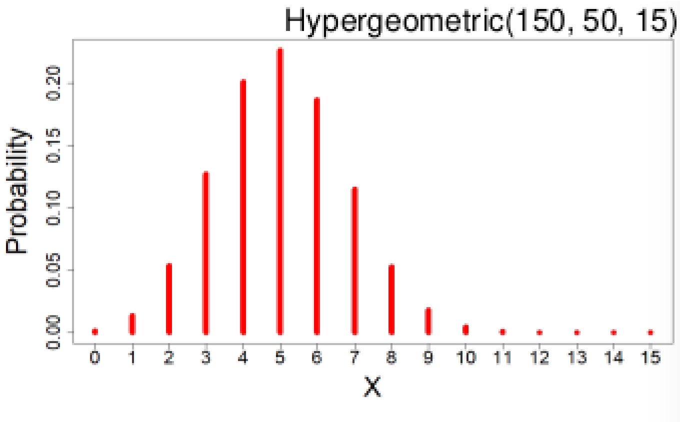
X ~ Geometric(p) □□ ∈ (0, 1]

If Y_1, Y_2, Y_3 ... are a sequence of independent Bernoulli(p) random variables then the number of failures before the first success, X, follows a Geometric distribution.

- PMF = P(X = x) = $p(1-p)^x$
- Expected value = E(X) = (1-p)/p
- Variance = V(X) = $(1-p)/p^2$
- MGF = $M_x(t) = p / (1 - (1 - p)e^t)$

Ex. We want to know the number of times to flip a coin before it lands on heads.

Hyper-Geometric Distribution



X ~ Hypergeometric(N, K, n)

Suppose a finite population of size N contains two mutually exclusive events: K success events and N-K failure events. If n events are randomly chosen *without* replacement X is the number of success events chosen.

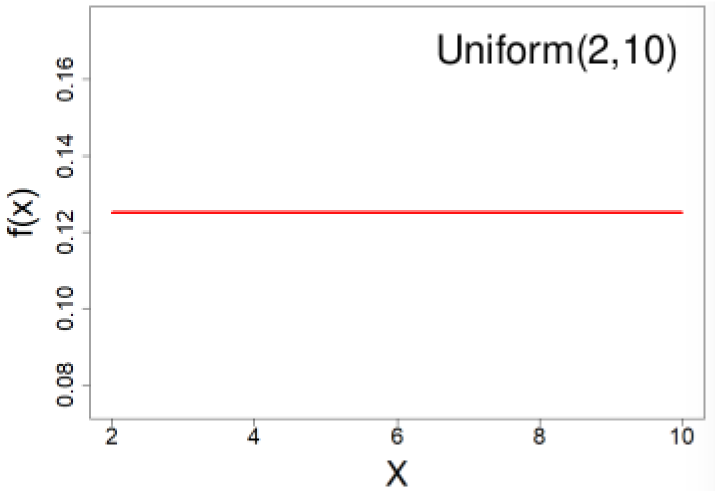
- PMF:

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

- Expected value = E(X) = nk / N
- Variance = V(X) = ((nK) / N) * ((N-K) / N) * ((N - n) / (N - 1))

Ex. A bag has 7 red beads and 13 white beads. If 5 are drawn *without* replacement what is the probability exactly 4 are red?

Uniform Distribution



All outcomes are equally likely, they can be discrete or continuous.

X ~ Uniform(a, b)   a < b

- PDF:

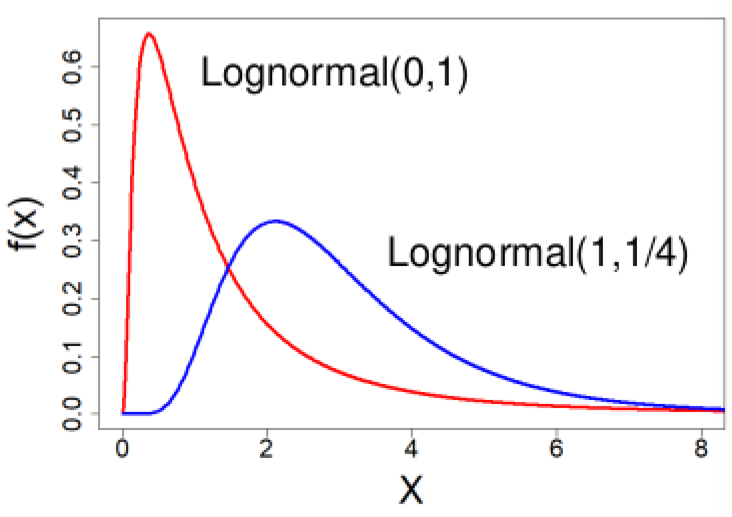
$$f_X(x) = \begin{cases} \frac{1}{b-a}, a \leq x \leq b \\ 0, \qquad otherwise \end{cases}$$

- E(X) = (a + b)/2
- V(X) = $(b - a)^2$ / 12
- CDF = F(X) = (x - a) / (b -a), a<=x<=b
- MGF:

$$M_X(t) = \begin{cases} \frac{e^{tb} - e^{ta}}{t(b-a)}, t \neq 0 \\ 1,\ t = 0 \end{cases}$$

We use this distribution we use when we have no idea how the data is distributed.

Log-Normal Distribution



X ~ Lognormal(μ , σ^2) -infinity < μ < infinity, σ > 0

- PDF:

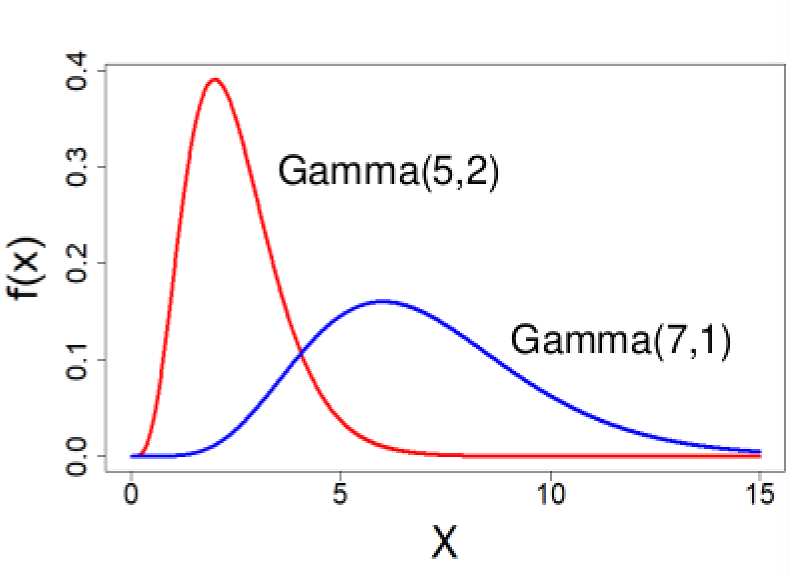
$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{\log(x) - \mu}{\sigma}\right)^2\right] \text{ for } x > 0$$

- E(X) = exp(μ + σ^2/2)
- Median = e^{μ}
- V(X) = μ^2 * ($e^{\sigma\text{^}2}$-1)

- $\log(X) \sim N(\mu, \sigma^2)$ - the log is normal
- These distributions are often skewed to the right

Ex. Amount of rainfall, production of milk by cows, or stock market fluctuation often follow logarithmic functions.

Gamma Distribution



$X \sim \text{Gamma}(\alpha, \lambda)$ $\alpha > 0$, $\lambda > 0$

Used to predict the wait time until the first of event of something.

- PDF

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \text{ for } x > 0$$

Alternate paramterization with α > 0, □□ = 1 / λ > 0 is used by R:

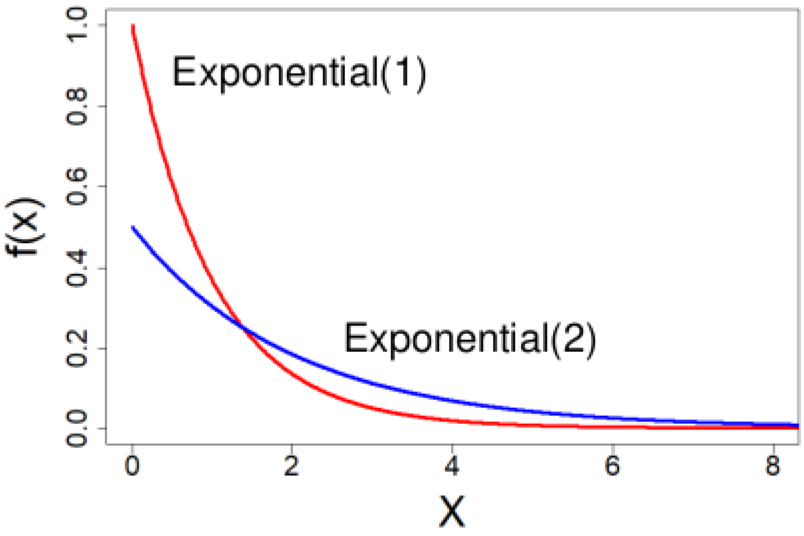
$$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

- $E(X) = \alpha / \lambda$
- $V(X) = \alpha / \lambda^2$
- MGF:

  $\left(\frac{\lambda}{\lambda - t}\right)^\alpha$, $t < \lambda$

Ex. Used to model time to failure or time to death.

Exponential Distribution



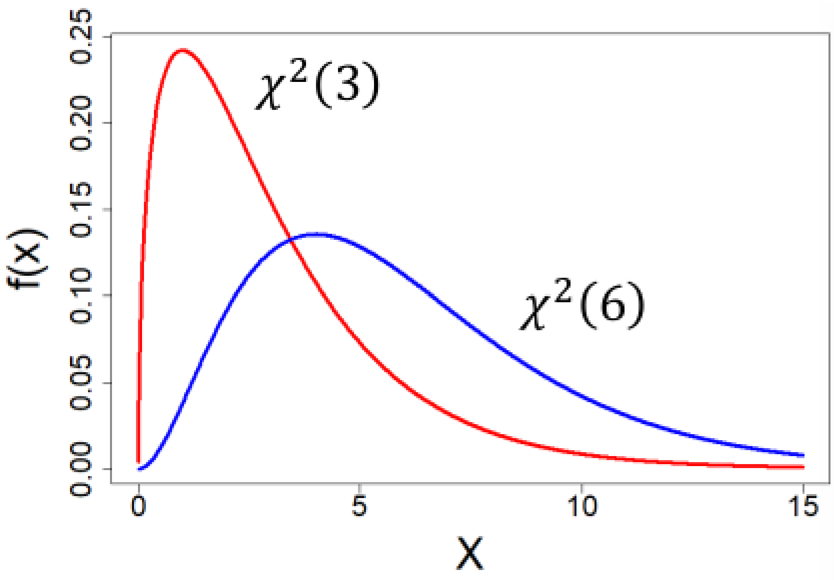
A special subset of the Gamma Distribution (α = 1)

X ~ Exponential(λ)   λ > 0

- PDF = $f_x(x) = \lambda e^{-\lambda x}$ for x > 0
- $E(X) = 1 / \lambda$
- $V(X) = 1 / \lambda^2$
- CDF = $F_x(x) = 1 - e^{-\lambda x}$
- MGF = $M_x(t) = \lambda / (\lambda - t)$, $t < \lambda$

Ex. The time between geyser eruptions.

Chi-Square Distribution



Special case of the Gamma Distribution (α = k/2, λ = 1/2)

X ~ X^2(k)   k is a positive integer (degrees of freedom, "df")

- PDF:

$$f_X(x) = \frac{1}{2^{k/2}\Gamma\left(k/2\right)} x^{(k/2)-1} e^{-x/2} \quad \text{for } x > 0$$

- E(X) = k
- V(X) = 2k
- MGF = $(1 - 2t)^{-k/2}$, t < 1/2

If you took a sample of Z scores and squared them you would have a chi-squared distribution with k = 1. Meaning, if Z_1, Z_2... Z_m are independent standard normal random variables then:

$$\sum_{i=1}^{m} Z_i^2 \sim \chi^2(m)$$

Very few real world distributions follow a chi-sqaure distribution, it is mainly used in hypothisis testing.

Bivariate Normal Distribution

A bivariate normal distribution is made up of two independent random variables. The two variables are both normally distributed, and have a normal distribution when added together.

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$$

σ_{12} = Cov(X_1, X_2)

PDF:

$$f(x_1, x_2) = \frac{\exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}))}{2\pi\sqrt{\det\boldsymbol{\Sigma}}}$$

Function of a Discrete Random Variable

Suppose X is a discrete random variable and Y is a function of X.   Y = g(X)

The Y is also a random variable: P(Y = y) = P(g(X) = y)

Example:

$$X = \begin{cases} 0, & \text{with probability } 1/3 \\ 1, & \text{with probability } 2/3 \end{cases}$$

$$Y = aX + b \quad \text{(a linear transformation)}$$

Then

$$Y = \begin{cases} b, & \text{with probability } 1/3 \\ a + b, & \text{with probability } 2/3 \end{cases}$$

Function of a Continuous Random Variable

Using the same equation as above but assuming the variables are coninuous random variables:

The PDF = $\frac{d}{dy} F_Y(y)$

The CDF =
$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P(\{x: g(x) \leq y\}) \end{aligned}$$

If g is one-to-one (strictly increasing or decreasing) then g has an inverse g^{-1}, in the above case:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

Properties of Expectation and Variance

$$E(a + bX) = a + bE(X)$$

$$E\big(g(X)\big) = \sum_i g(x_i)p_X(x_i) \text{ or } \int g(x)f_X(x)dx$$

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i) \quad \text{if the } X_i's \text{ are independent}$$

$$V(a + bX) = b^2 V(X)$$

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) \quad \text{if the } X_i's \text{ are independent}$$

Discrete Multivariate Distributions

Let X and Y be discrete random variables

- Joint probability mass function
  - $0 \leq p_{X,Y}(x, y) \leq 1$
  - $\sum_{x,y} p_{X,Y}(x, y) = 1$
- Marginal probability mass functions
  - $p_X(x) = \sum_y p_{X,Y}(x, y)$ sum $p_{X,Y}(x, y)$ for all values of y
  - $p_Y(y) = \sum_x p_{X,Y}(x, y)$ sum $p_{X,Y}(x, y)$ for all values of x
- Conditional probability mass functions
  - $p_{X|Y}(x|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$
  - $p_{Y|X}(y|X = x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$
- X and Y are independent if and only if
  - $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ for all x and y

Continuous Multivariate Distributions

Let X and Y be continuous random variables

- Joint density function
  - $f_{X,Y}(x,y) \geq 0$
  - $\iint f_{X,Y}(x,y)dxdy = 1$
- Marginal density functions
  - $f_X(x) = \int f_{X,Y}(x,y)dy$
  - $f_Y(y) = \int f_{X,Y}(x,y)dx$
- Conditional density functions
  - $f_{X|Y}(x|Y=y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
  - $f_{Y|X}(y|X=x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
- X and Y are independent if and only if
  - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x and y

Covariance and Correlation

Correlation is defined as an indication as to how strong the relationship between the two variables is:

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}}, \qquad -1 \leq \rho_{X,Y} \leq 1$$

A positive correlation has σ > 0 and negative correlation has σ < 0

Covariance provides information about how the variables vary together:

cov(X, Y) = R[(X - E(X))(Y - E(Y))]

This is also equivalent to:

cov(X, Y) = E(XY) - E(X)*E(Y)

Thus if X and Y are independent:

cov(X, Y) = corr(X, Y) = 0

However cov(X, Y) = 0 does not imply indepence unless they are jointly normally distributed.

Conditional Expectation of X given Y = y, denoted E(X | Y = y):

Discrete random variables

- $E(X|Y=y) = \sum_{i=1}^{K} x_i p_{X|Y}(x_i|y)$ use the conditional PMF

Continuous random variables

- $E(X|Y=y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx$ use the conditional PDF

Conditional variance can be defined similarly (use the conditional PMF or PDF)

Revision #9
Created 19 August 2022 14:04:38 by Elkip
Updated 20 August 2022 16:14:55 by Elkip