Probability Theory Lecture 09

Padmini Mukkamala

Budapest University of Technology and Economics

January 21, 2021

Overview

1. Normal Distribution
2. Bivariate Normal Distribution

Normal Distribution - what's in a name?

The normal distribution was introduced by the French mathematician Abraham DeMoivre in 1733. DeMoivre, who used this distribution to approximate probabilities connected with coin tossing, called it the exponential bell-shaped curve. Its usefulness, however, became truly apparent only in 1809, when the famous German mathematician Karl Friedrich Gauss used it as an integral part of his approach to predicting the location of astronomical entities. As a result, it became common after this time to call it the Gaussian distribution.

Normal Distribution - what's in a name?

During the mid- to late 19th century, however, most statisticians started to believe that the majority of data sets would have histograms conforming to the Gaussian bell-shaped form. Indeed, it came to be accepted that it was 'normal' for any well-behaved data set to follow this curve. As a result, following the lead of the British statistician Karl Pearson, people began referring to the Gaussian curve by calling it simply the normal curve. – source Sheldon Ross, page 207.

Normal Distribution

Definition

A continuous random variable X is said to have Normal distribution $N(\mu, \sigma^2)$, if its probability density function is defined as follows:

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

Where $\mu \in \mathcal{R}$ and $\sigma \in [0, \infty)$ or $\mathcal{R}^+$.

To see that this is a valid probability distribution, we need that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = 1$$

Proof that the pdf is valid

Proof: We will substitute $z = \frac{x-\mu}{\sigma}$ to simplify our integral. So we will need to show that, $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$
Let $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$. Then $I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dxdy$.
Here we will use polar coordinates for solving the double integral.
We note that $x = rcos\theta$ and $y = rsin\theta$ and $dxdy = rdrd\theta$. Using these substitutions, we get $\int_0^{\infty} \int_0^{2\pi} \frac{1}{2\pi} e^{-\frac{r^2}{2}} rdrd\theta$. This we can easily evaluate to get 1.

Standard Normal Distribution

When the parameters $\mu = 0$ and $\sigma^2 = 1$, then the distribution is called the **Standard Normal Distribution** and denoted by $N(0,1)$. We usually use the letter Z for a random variable with standard normal distribution. So,

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Standard Normal Distribution

CDF of the Standard Normal Distribution

The CDF of the standard Normal distribution is denoted by $\Phi(a)$ (pronounced as Fi in Five). So, for $Z \sim N(0,1)$, $\Phi(a) = P(Z \leq a) = \int_{-\infty}^{a} \frac{1}{2\pi} e^{-\frac{1}{2}z^2} dz$. This integral doesn't have a nice solution in elementary functions but can be approximated.

The density function $f_Z(z)$, or the bell curve, is symmetric about 0. This gives us the very useful property that $P(Z \leq -a) = P(Z \geq a), \forall a$.

Standard Normal Distribution - CDF Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Standard Normal Distribution

Use the table to find the following probabilities:

- $P(Z < 1)$
- $P(Z < -1)$
- $P(Z > 1)$
- $P(-1 < Z < 1)$

Transformation of a Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Further, let $Y = aX + b$. What will be the distribution of Y?

Transformation of a Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Further, let $Y = aX + b$. What will be the distribution of Y?

$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a})$

Differentiating with respect to y,

$f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a}) = \frac{1}{\sqrt{2\pi}\sigma a} e^{-\frac{1}{2}\frac{(y-(a\mu+b))^2}{a^2\sigma^2}}$.

Transformation of a Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Further, let $Y = aX + b$. What will be the distribution of Y?

$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a})$

Differentiating with respect to y,

$f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a}) = \frac{1}{\sqrt{2\pi}\sigma a} e^{-\frac{1}{2}\frac{(y-(a\mu+b))^2}{a^2\sigma^2}}$.

So $Y \sim N(a\mu + b, a^2\sigma^2)$. In particular, any linear transformation of a Normal distribution is also normal.

Transformation of a Normal Distribution cont.

Taking $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$ above, we get $Y \sim N(0, 1)$. This method of linearly transforming a random variable with a general Normal distribution to a random variable with the Standard Normal Variable is called **Standardization**.

Further note that

$F_X(x) = P(X \leq x) = P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) = F_Y(\frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma})$.

General Normal Distribution

In a study of half-giants and their heights, Prof. Dumbledore found that their mean height was 3m, with standard deviation 1m. If you let X be a random variable denoting a half-giant's height, find the following probabilities:

- $P(X < 4)$
- $P(X < 2)$
- $P(X > 4)$
- $P(2 < X < 4)$

Moments of the Standard Normal Distribution

Let $M_i = E(Z^i)$ denote the i^{th} moment of the standard normal distribution $N(0,1)$. Then,

$M_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^i e^{-\frac{1}{2}z^2} dz$.

We notice here that if we set $u = z^{i-1}$ and let $dv = ze^{-\frac{1}{2}z^2} dz$, then $v = -e^{-\frac{1}{2}z^2}$.

Using integration by parts, $\int udv = uv - \int vdu$, we get,

$M_i = [\frac{-z^{i-1}e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (i-1)z^{i-2}e^{-\frac{1}{2}z^2} dz$.

Moments of the Standard Normal Distribution cont.

We can see that the first term evaluates to 0, while the second term is a smaller moment. So we get the following recurrance:

Recurrence for Moments of Z, Standard Normal RV

$M_i = (i-1)M_{i-2}$.

So if we find the moments M_0 and M_1, then we can find all moments of the Standard Normal Distribution.

Odd Moments and Expected Value

The **Expected value**, or the First Moment of the Standard Normal Variable Z is

$$\int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Using the substitution that $z^2 = y$, we can check that this is 0. If $X \sim N(\mu, \sigma^2)$, then since $X = \sigma Z + \mu$, $E(X) = \mu$.

Odd Moments and Expected Value

First Moment

$$E(Z) = 0,\ \ E(X) = \mu$$

Odd Moments of Z

From this, we can conclude that odd moments are zero.

$$E(Z^{2i+1}) = 0, \forall i \in \mathcal{N}$$

Even Moments

We know that $M_0 = 1$, because, $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$
Then $M_2 = E(Z^2) = (2-1)M_0 = 1$.
So the $Var(Z) = E(Z^2) = 1$.
Further even moments are given by $M_{2i} = (2i-1)(2i-3)...5.3.1$.

Even Moments

We know that $M_0 = 1$, because, $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$
Then $M_2 = E(Z^2) = (2-1)M_0 = 1$.
So the $Var(Z) = E(Z^2) = 1$.
Further even moments are given by $M_{2i} = (2i-1)(2i-3)...5.3.1$.
If $X \sim N(\mu, \sigma^2)$, then $X = \sigma Z + \mu$ and $Var(X) = \sigma^2 Var(Z) = \sigma^2$.
And so, the standard deviation of X, $\sigma_X = \sigma$.

Even Moments

Even Moments

Even moments of the standard normal random variable Z,

$$E(Z^{2i}) = M_{2i} = (2i-1)(2i-3)...5.3.1$$

Variance and Standard Deviation

For $Z \sim N(0,1)$, the standard normal variable,

$$E(Z) = 0, \text{ and } Var(Z) = \sigma_Z = 1$$

For $X \sim N(\mu, \sigma^2)$, a variable with general normal distribution,

$$E(X) = \mu, Var(X) = \sigma^2 \text{ and } \sigma_X = \sigma$$

Moment Generating Function

The Moment Generating Function of the Standard Normal Distribution $N(0,1)$ is given by

$$E(e^{Zt}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{zt} e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2zt)} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2zt+t^2)+\frac{1}{2}t^2} dz$$

$$= e^{\frac{1}{2}t^2}$$

Moment Generating Function

The Moment Generating Function of the Standard Normal Distribution $N(0,1)$ is given by

$$E(e^{Zt}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{zt} e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2zt)} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2 - 2zt + t^2) + \frac{1}{2}t^2} dz$$

$$= e^{\frac{1}{2}t^2}$$

Using Linearity and Scaling properties of MGFs, we can see that the MGF of the general normal distribution is $= e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

Summary

- Density Function of Normal distribution:
  $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$
- Standardization $Z = \frac{X-\mu}{\sigma}$
- CDF of $N(0,1)$, the Standard Normal Distribution, Φ is given in table.
- $P(X \leq x) = \Phi(\frac{x-\mu}{\sigma})$.
- For $Z \sim N(0,1)$, $E(Z^i) = (i-1)E(Z^{i-2})$.
- $E(Z) = 0,\ E(X) = \mu,\ Var(Z) = 1,\ Var(X) = \sigma^2$.
- MGF of $N(0,1)$ is $e^{\frac{t^2}{2}}$, and of $N(\mu,\sigma^2)$ is $e^{\mu t+\frac{\sigma^2 t^2}{2}}$.

Standard Bivariate Normal distribution

Denoted by $\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, the standard bivariate normal random variable consists of two **independent** random variables with standard normal distribution. So, the joint probability density function is given by,

$$f_{Z_1,Z_2}(z_1, z_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2+z_2^2)}$$

Standard Bivariate Normal distribution

Properties

$$E(\boldsymbol{Z}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The covariance matrix is the identity matrix. Recall,

$$Cov(\boldsymbol{Z}) = \begin{pmatrix} Var(Z_1) & Cov(Z_1, Z_2) \\ Cov(Z_1, Z_2) & Var(Z_2) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Convolution of Normals

Theorem

Given two independent random variables X, Y with normal distribution, their linear combination is also Normal.
More mathematically, let $X_i \sim N(\mu_i, \sigma_i^2)$, for $i = 1, 2$ be independent. Then, for any two non-zero constants $c_1, c_2 \in \mathcal{R}$, $c_1 X_1 + c_2 X_2$ has the normal distribution $N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2^2\sigma_2^2)$.
Proof: use Moment Generating Function!

Bivariate Normal distribution

Recall that in the 1D case, the general normal distribution was a linear transform of the standard normal. That is $X = \sigma Z + \mu$.

Bivariate Normal distribution

Recall that in the 1D case, the general normal distribution was a linear transform of the standard normal. That is $X = \sigma Z + \mu$.
A 2D random variable $\boldsymbol{X}$ has a bivariate normal distribution if $\exists \boldsymbol{A} \in \mathcal{R}^{2\times 2}$ and a $\boldsymbol{\mu} \in \mathcal{R}^2$ such that $\boldsymbol{X} = \boldsymbol{AZ} + \boldsymbol{\mu}$. (So every bivariate normal distribution is obtained as a Linear combination of standard bivariate normal distribution).

Bivariate Normal distribution

Properties

$$E(\boldsymbol{X}) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \boldsymbol{\mu}$$

Bivariate Normal distribution

Properties

$E(\boldsymbol{X}) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \boldsymbol{\mu}$

Recall that the covariance matrix for a linear transform $\boldsymbol{AX}$ is given by $\boldsymbol{A}\Sigma\boldsymbol{A}^T$. Since $\boldsymbol{\Sigma}$ here is the Covariance matrix of the standard bivariate normal distribution, which is the Identity matrix, so it follows that $Cov(\boldsymbol{X}) = \boldsymbol{AA}^T$.

Joint Probability Density Function

Recall, in the one dimensional case, the pdf for a normal distribution was as follows:

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

Notice that we can write this as:

$f_X(x) = \frac{1}{\sqrt{2\pi}|\sigma|} e^{-\frac{1}{2}(x-\mu)(Var(X))^{-1}(x-\mu)}$.

We see that the joint pdf of the bivariate general normal distribution is given by:

$f_{\boldsymbol{X}}(x_1, x_2) = \frac{1}{(\sqrt{2\pi})^2|\Sigma|} e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T(\boldsymbol{\Sigma})^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}$.

Example

For a random variable vector, $\boldsymbol{X}$, we know that

$\boldsymbol{X} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \boldsymbol{Z} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, where $\boldsymbol{Z}$ is the standard bivariate normal distribution vector. Find the Covariance matrix and pdf of $\boldsymbol{X}$.

Example

Covariance matrix is $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

We can see that $det(\boldsymbol{\Sigma}) = 1$ and $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$.

Plugging it into the joint pdf formula we get,

$f_{\boldsymbol{X}}(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}((x-2)^2 - 4(x-2)(y-1) + 5(y-1)^2}$

$= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + 5y^2 - 4xy - 2y + 1)}$.

Properties

The components X_1, X_2 of a bivariate normal distribution have some nice properties. **Important:** These properties don't hold for two general variables X, Y with normal distribution, but only if we know that together they have a joint bivariate normal distribution.

- Linear combination $c_1X_1 + c_2X_2$, for non-zero c_1, c_2, is Normal.
- If X_1, X_2 are uncorrelated, then they are independent. (Counter example to this in general case is let X be standard normal and $Y = WX$ where W takes values 1 and -1 with probability $\frac{1}{2}$.

Properties

The components X_1, X_2 of a bivariate normal distribution have some nice properties. **Important:** These properties don't hold for two general variables X, Y with normal distribution, but only if we know that together they have a joint bivariate normal distribution.

- Linear combination $c_1X_1 + c_2X_2$, for non-zero c_1, c_2, is Normal.
- If X_1, X_2 are uncorrelated, then they are independent. (Counter example to this in general case is let X be standard normal and $Y = WX$ where W takes values 1 and -1 with probability $\frac{1}{2}$.
  Proof: The joint pdf splits neatly into the product of two.
- Regression $E(X_2|X_1)$ is the linear regression. (Later!)

Bell curve

Mathematicians for decades have marvelled at the beauty of a normal distribution. To quote,
I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error. " The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. – Sir Francis Galton.

Radially symmetric distributions

Suppose X, Y are two **independent** random variables, where we also know that the joint density function only depends on the distance from the origin. What can we say about $f_{X,Y}$?

Radially symmetric distributions

Suppose X, Y are two **independent** random variables, where we also know that the joint density function only depends on the distance from the origin. What can we say about $f_{X,Y}$?
So we know, $f_{X,Y}(x, y) = g(x^2 + y^2)$ where g is any function.
Because of independence, we also know that
$f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Radially symmetric distributions cont.

We can partially differentiate with respect to x, the equation, $f_X(x)f_Y(y) = g(x^2 + y^2)$, to get $f'{}_X(x)f_Y(y) = 2xg'(x^2 + y^2)$.
If we divide these two equations, we get, $\frac{f'_X(x)}{2xf_X(x)} = \frac{g'(x^2+y^2)}{g(x^2+y^2)}$
Since the right hand is the same for any two points at the same distance from the origin, we can see that it must be the constant function. So,
$\frac{f'_X(x)}{f_X(x)} = c.2x$

Radially symmetric distributions cont.

Integrating this, we get,
$ln(f_X(x)) = cx^2 + d$, where $c, d \in \mathcal{R}$.
or, $f_X(x) = ae^{cx^2}$, where $c \in \mathcal{R}$ and $a \in \mathcal{R}^+$.
For the condition that $\int_{-\infty}^{\infty} fdx = 1$, we can compute the constants to get that function is $N(0, \sigma^2)$ for some $\sigma > 0$.

The End