# Probability and Statistics: From Classical to Bayesian
# Week 4

13th May - 20th May 2021

## 1 Change of Variable in $\mathbb{R}^m$

When we needed to find find probability distributions of functions of a single variable we worked with normal pdfs and pmfs, For example when we had to know the distribution of $X^2$,$|X|$ and so on. What if we were working with two variables? Lets say I knew about X and Y and was interested in knowing the distribution of XY, just knowing about the distributions of X and Y doesnt suffices. How does X and Y behave with each other is also important, this is where Joint pdfs come into play, lets see how we work with them.

**CDF technique** for this is pretty similar to that of monovariate, and now that you a know bit about Joint Distributions, you will get enough feel for it through an example. That'd be given in adjoining PDF.

### 1.1 PDF technique for continuous distributions

The idea is quite similar to the univariate case, and will be discussed for a special class of functions.
Let $X = (X_1, \ldots, X_m)$ be a random vector with density $f(t_1, \ldots, t_m)$. Let $T : \mathbb{R}^m \to \mathbb{R}^m$ be a one-one function which is continuously differentiable.
Let $\mathbf{Y} = \mathbf{T}$(X). In co-ordinates, we may write $Y = (Y_1, \ldots, Y_m)$ and $Y_1 = T_1(X_1, \ldots, X_m), \ldots, Y_m = T_m(X_1, \ldots, X_m),$ where $T_i : \mathbb{R}^m \to \mathbb{R}$ are the components of $T$.

**The change of variable formula**: In the setting described above, the joint density of $Y_1, \ldots, Y_m$ is given by

$$g(y) = f(T^{-1}y)|J[T^{-1}]y|$$

where $|J[T^{-1}](y)|$ is the Jacobian determinant (recall from MTH101) of the function $T^{-1}$ at the point $y = (y_1, \ldots, y_m)$ You will get to know its utility and usage better with examples in the adjoining PDF.
Notice the similarity with the univariate case.

## 2 Independent and Identically distributed (i.i.d)

**Identically Distributed Random Variables**: Two random variables X and Y are said to have the same distribution (written as X $\stackrel{D}{=}$ Y ) if they have the same distribution function, i.e., if

$$F_X(x) = F_Y(x) \ \forall x \in R$$

Which is equivalent to saying that X and Y have the same probability distributions which is equivalent to saying that they have the same moment generating function. i.e.

$$M_X(t) = M_y(t) <=> F_X(x) = F_Y(x)$$

**Independence of variables**: Let $X = (X_1, ..., X_m)$ be a random vector (this means that $X_i$ are random variables on a common probability space). We say that $X_i$s are independent if

$$F_X(t_1, ..., t_m) = F1(t_1)...Fm(t_m) \ \forall \ t_1, ..., t_m.$$

Remark. In case $X_1, .., X_m$ have a joint pmf or a joint pdf (which we denote by $f(t_1, ..., t_m)$), the condition for independence is equivalent to

$$f(t_1, ..., t_m) = f_1(t_1)...f_m(t_m)$$

where $f_i$ is the marginal density (or pmf) of $X_i$. This fact can be derived from the definition easily.
For example, in the case of densities, observe that

$$\begin{aligned} f(t_1, ..., t_m) &= \frac{d^m}{dt_1 dt_2 ... dt_m} F(t_1, ..., t_m) \text{ (true for any joint density)} \\ &= \frac{d^m}{dt_1 dt_2 ... dt_m} F_1(t_1)...F_m(t_m) \text{ (by independence)} \\ &= f_1(t_1)...f_m(t_m) \end{aligned}$$

When we turn it around, this gives us a quicker way to check independence.
Fact: Let $X_1, ..., X_m$ be random variables with joint pdf $f(t_1, ..., t_m)$. Suppose we can write this pdf as:

$$f(t_1, ..., t_m) = cg_1(t_1)...g_m(t_m)$$

where c is a constant and $g_i$ are some functions of one-variable. Then, $X_1, ..., X_m$ are independent. Further, the marginal density of $X_k$ is $c_k g_k(t)$, where $c_k$ is the normalizing constant( $\frac{1}{\int_{-\infty}^{\infty} g_k(s)ds}$).

## 3 Conditional Probabilities using Joint pdf for Random Variables

We assume a set of Variables, say $X = (X_1, \ldots, X_{k+l})$. Let $f(t_1, \ldots, t_{k+l})$ be the pmf of $(X_1, \ldots, X_{k+l})$ and let $g(t_1, \ldots, t_l)$ be the pmf of $(X_{k+1}, \ldots, X_{k+l})$ (of course we can compute $g$ from $f$ by summing over the first k indices). Then, for any $s_1, \ldots, s_l$ such that $P\{X_{k+1} = s_1, \ldots, X_{k+l} = s_l\} > 0$, we can define

$$h_{s_1,\ldots,s_l}(t_1, \ldots, t_k) = P\{X_1 = t_1, \ldots, X_k = t_k | X_{k+1} = s_1, \ldots, X_{k+l} = s_l\} = \frac{f(t_1, ..., t_k, s_1, ..., s_l)}{g(s_1, ..., s_l)}$$

$h_{s_1,\ldots,s_l}$ is called the conditional pmf of $(X_1,\ldots,X_k)$ given that $X_{k+1} = s_1,\ldots,X_{k+l} = s_l$.

Above is for Random Variables with PMF, but this is for you to get better understanding about conditional probability. For non-trivial cases, this makes sense when Random Variables are not independent.

This section would be used a lot in Bayesian part of the course. For example, let $f(X,\theta|\text{data})$ be joint pdf of $X$ and $\theta$,given data(if you don't get it right now, its okay. This will be explained in good detail again during Bayesian Analysis), where each of them can be a vector.

Then for getting pdf of $\theta$ given X,data can be thought of as :

$$f(X,\theta|data) = f(\theta|X,data)f(X|data)$$

This is generalised Bayes Rule, used on pdfs (and Conditional Probability). $f(X|data)$ can be calculated by integrating out $\theta$ component from $f(X,theta|data)$. You learnt this last week under Joint pdfs.

An example for this would be given in the adjoining PDF.

# 4 Moments and MGFs

This part will be heavy in formulae and standard terminologies, the main takeaway is how to write MGFs for joint distribution, how does independence affects expectations and the new term covariance.

- If X is of discrete type, then

$$E(\psi(X)) = \sum \psi(x)f_X(x)$$

- If X is of absolutely continuous type, then

$$E(\psi(X)) = \int \psi(x)fX(x)dx$$

- For non-negative integers $k_1,...,k_p$. let $\psi(x) = x_1^{k_1}...x_p^{k_p}$. Then,

$$\mu'_{k_1,...,k_p} = E[\psi(X)]$$

  is called a joint raw moment of order $k_1 + ... + k_p$ of X.

- For non-negative integers $k_1,...,k_p$. let $\psi(x) = (x_1 - E[X_1])^{k_1}...(x_p - E[X_p])^{k_p}$. Then,

$$\mu_{k_1,...,k_p} = E[\psi(X)]$$

  is called a joint central moment of order $k_1 + ... + k_p$ of X.

- Let $\psi(x) = (x_i - E(X_i))(x_j - E(X_j))$ for i,j = 1,..,p. Then, the covariance between $X_i$ and $X_j$ is

$$Cov(X_i,X_j) = E((X_i - (X_i))(X_j - E(X_j))) = E[X_iX_j] - E[X_i]E[X_j]$$

Let $X = (X_1,X_2,...,X_{p_1})$ and $Y = (Y_1,Y_2,...,Y_{p_2})$ be random vectors, and let $a_1,...,a_{p_1}$ and $b_1,...,b_{p_2}$ be real constants. Assume that the involved expectations exist. Then,

- $E(\sum_{i=1}^{p_1} a_i X_i) = \sum_{i=1}^{p_1} a_i E(X_i)$
- $Cov(\sum_{i=1}^{p_1} a_i X_i, \sum_{j=1}^{p_2} b_i Y_i) = \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} a_i b_j Cov(X_i, Y_j)$

In particular,

$$Var(\sum_{i=1}^{p_1} a_i X_i) = \sum_{i=1}^{p_1} a_i^2 Var(X_i) + 2 \sum_{1 \le i < j \le p_1} a_i a_j Cov(X_i, X_j)$$

**Moment Generating Functions:**

$$M_X(t) = E(e^{\sum_{i=1}^{p} t_i X_i}) \text{ where } t = (t_1, t_2, ..., t_p)$$

As in the one-dimensional case, many properties of probability distribution of X can be studied through the joint mgf of X. Some of the results, which may be useful in this direction, are provided below (without their proofs). Note that $M_X(0_p) = 1$, where $0_p$ is the vector of 0s. If $X_1, ..., X_p$ are independent, then

$$M_X(t) = E(e^{\sum_{i=1}^{p} t_i X_i}) = E(\prod_{i=1}^{p} e^{t_i X_i}) = \prod_{i=1}^{p} E(e^{t_i X_i}) = \prod_{i=1}^{p} M_{X_i}(t_i)$$

Suppose that $M_X(t)$ exists in a rectangle $(-a, a) \in R^p$. Then, $M_X(t)$ possesses partial derivatives of all orders in (-a, a). Furthermore, for positive integers $k_1, \ldots, k_p$

$$E[X_1^{k_1} X_2^{k_2} \ldots X_p^{k_p}] = \left[ \frac{d^{k_1+k_2+\cdots+k_p}}{dt_1^{k_1} \ldots dt_p^{k_p}} M_{\mathbf{X}}(\mathbf{t}) \right]_{t=\mathbf{0}_p}$$

For i $\neq$ j with i, j $\in \{1, \ldots, p\}$, define

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$
$$= \left[ \frac{d^2}{dt_i t_j} M_X(t) \right]_{t=\mathbf{0}_p} - \left[ \frac{d}{dt_i} M_X(t) \right]_{t=\mathbf{0}_p} \left[ \frac{d}{dt_j} M_X(t) \right]_{t=\mathbf{0}_p}$$
$$= \left[ \frac{d^2}{dt_i dt_j} \psi_X(t) \right]_{t=\mathbf{0}_p}$$

where $\psi_X(t) = lnM_X(t)$.

If $X_1, X_2, ..., X_p$ are independent and identically distributed (i.i.d.), i.e., $X_i \stackrel{D}{=} X_1$ for i = 2,..,p, then

$$M_X(t) = \prod_{i=1}^{p} M_{X_1}(t_i) \text{ for } t \in R^p$$

Define $Y = \sum_{i=1}^{p} X_i$ and $\bar{X} = Y/p$, then

$$M_Y(t) = [M_{X_1}(t)]^p \text{ and } M_{\bar{X}}(t) = [M_{X_1}(t/p)]^p \text{ for } t \in R$$

**Exercise**. Try finding expectation of Y (E[Y]), in terms of the expectation of $X_i's$, similarly for $\bar{X}$. (Hint: Use their MGFs that are written above).

**Correlation**: Let X,Y be random variables on a common probability space. Their correlation is defined as

$$Corr(X,Y) = \frac{Cov(X,Y)}{Var(X)Var(Y)}$$

**Measures of association**: The marginal distributions of X and Y do not determine the joint distribution of (X,Y). In particular, giving the means and standard deviations of X and Y does not tell anything about possible relationships between the two.

Covariance is the quantity that is used to measure the "association" of Y and X. Correlation is a dimension free quantity that measures the same. For example, we shall see that if Y = X, then Corr(X,Y) = +1, if Y = -X then Corr(X,Y) = -1. Further, if X and Y are independent, then Corr(X,Y) = 0. In general, if an increase in X is likely to mean an increase in Y , then the correlation is positive and if an increase in X is likely to mean a decrease in Y then the correlation is negative.
Read this: `http://probability.ca/jeff/teaching/uncornor.html`.

# 5 Multivariate Normal Distribution

Some of the Proofs and Derivation for this part would require Matrix Algebra you'd learn as a part of MTH102. Since we do not presume anything about that knowledge, we'd just be quoting result. Anyone interested is welcomed to clear it on Discord channel.

## 5.1 Bivariate Normal

We say that (X,Y) follows a bivariate normal distribution if its pdf is given by

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-Q/2} \quad -\infty < x < \infty \quad -\infty < y < \infty$$

where

$$Q = \frac{1}{1-\rho^2}\left[(\frac{x-\mu_1}{\sigma_1})^2 - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2}) + (\frac{y-\mu_2}{\sigma_2})^2\right]$$

with $-\infty < \mu_i < \infty$; $\sigma_i > 0$ for i = 1,2 and $\rho$ satisfies $\rho^2 < 1$. Clearly, this function is positive everywhere in $R^2$.

There are other ways to write the pdf as well, we think this is one of the better ways since it closely resembles the Univariate case and the constants here store useful information about the pdf. We will see this soon.

MGF for Bivariate Normal Distribution**(X,Y)** is:

$$M_{(X,Y)}(t_1,t_2) = \exp\{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(t_1^2\sigma_1^2 + 2t_1t_2\rho\sigma_1\sigma_2 + t_2^2\sigma_2^2)\}$$

MGF for Univariate Normal Distribution X is:

$$M_X(t_1) = M_{(X,Y)}(t_1, 0) = \exp\{t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2\}$$

Notice that the MGF of X directy tells us that it follows a normal distribution. You can see that $\mu_1$ is mean of X and $\sigma_1$ is Var(X) and similarly for Y. Also you can check that $Cov(X, Y) = \rho\sigma_1\sigma_2$. Which gives us $\rho$ as the Correlation of X and Y.

## 5.2 Multivariate Normal

For Multivariate case of $\mathbf{X}(X_1, \ldots, X_m)$

$$M_{\mathbf{X}}(\mathbf{t}) = \exp\{\mathbf{t}'\mu + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\}$$

where $\mathbf{t}$ is $m \times 1$ vector, $\mu$ is $m \times 1$ vector having respective means as entry, $\Sigma$ is $m \times m$ matrix with $i, j^{th}$ entry being $Cov(X_i, X_j)$. $X_i$ are each random variables with normal distributions with mean $\mu_i$ and variance $\Sigma_{(i,i)}$.

**Linear Transformations in Multivariate Normal Distribution**: Suppose X has a $N_n(\mu, \Sigma)$ distribution. Let $Y = AX + \mathbf{b}$, where A is an $m \times n$ matrix and b $\in \mathbb{R}^m$. Then, Y has a $N_m(A\mu + \mathbf{b}, A\Sigma A')$ distribution. This can be shown by calcualting the MGF of Y, using the previous result that we gave for MGF of X.