University of California, Los Angeles
Department of Statistics

Statistics 403 **Instructor: Nicolas Christou**

Review of basic concepts

- Discrete random variables:
  Consider a random experiment with a sample space $\mathcal{C}$. A discrete random variable X is a function which assigns to each element $c \in \mathcal{C}$ one real value $X(c) = x$. The space of X, $\mathcal{A}$, is the set of real values such that $X(c) = x$.

- Probability mass function of a discrete random variable:
  $p(x) = P(X = x), \forall\ x \in \mathcal{A}$.
  $p(x) = 0, \forall\ x \notin \mathcal{A}$.
  $\Sigma_{\mathcal{A}}\, p(x) = 1$.

- Common discrete distributions:
  - Bernoulli
  - Binomial
  - Negative binomial
  - Hypergeometric
  - Poisson
  - Geometric

- Cumulative Distribution function (cdf):
  $F_X(x) = P(X \leq x) = \Sigma_{w \leq x}\, p(w)$, for discrete random variables.

- Continuous random variables:
  A random variable X is continuous if there exists a function $f(x) \geq 0$ such that

  $$P(X \in A) = \int_A f(x)dx.$$

  We call $f(x)$ the probability density function of X.

- Let X be a continuous random variable with range $\mathcal{A}$. Then,

  $$\int_{\mathcal{A}} f(x)dx = 1, \; f(x) = 0, \; \forall \; x \; \notin \mathcal{A}.$$

- Let X be a continuous random variable with pdf $f(x)$ and $-\infty < X < \infty$. Then

  $$\begin{aligned} P(X > a) &= \int_a^\infty f(x)dx \\ P(X < a) &= \int_{-\infty}^a f(x)dx \\ P(a < X < b) &= \int_a^b f(x)dx \end{aligned}$$

- Note that when X is a continuous random variable the following is true:

  $$P(X \geq a) = P(X > a)$$

  This is NOT true for discrete random variables.

- Cumulative distribution function (cdf):

  $$F(x) = P(X \leq x) = \int_{-\infty}^x f(x)dx,$$

  for continuous random variables.

- Therefore

  $$f(x) = \frac{\partial F(x)}{\partial x}$$

- Common continuous distributions:
  - Normal
  - Uniform
  - Gamma
  - Exponential
  - Beta
  - Chi-squared (central and non-central)
  - t (central and non-central)
  - F (central and non-central)

Explore relationships between distributions!

Probability distributions - Summary

Discrete distributions				
Distribution	Probability Mass Function	Mean	Variance	MGF
Binomial	$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$ $x=0,1,\cdots,n$	np	$np(1-p)$	$[pe^t+(1-p)]^n$
Geometric	$P(X=x)=(1-p)^{x-1}p$ $x=1,2,\cdots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Negative Binomial	$P(X=x)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$ $x=r,r+1,\cdots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$[\frac{pe^t}{1-(1-p)e^t}]^r$
Hypergeometric	$P(X=x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ $x=0,1,\cdots,n$ if $n\leq r$, $x=0,1,\cdots,r$ if $n>r$	$\frac{nr}{N}$	$n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$	
Poisson	$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$ $x=0,1,\cdots$	λ	λ	$exp[\lambda(e^t-1)]$
Continuous distributions				
Distribution	Probability Density Function	Mean	Variance	MGF
Uniform	$f(x)=\frac{1}{b-a}$ $a\leq x\leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Gamma	$f(x)=\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}$, $\alpha,\beta>0$, $x\geq 0$	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}$
Exponential	$f(x)=\lambda e^{-\lambda x}$, $\lambda>0$, $x\geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$(1-\frac{1}{\lambda}t)^{-1}$
Beta	$f(x)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ $\alpha>0,\ \beta>0,\ 0\leq x\leq 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	
Normal	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ $-\infty<x<+\infty$	μ	σ^2	$e^{\mu t+\frac{t^2\sigma^2}{2}}$

Exponential families

A probability density function or probability mass function is called an exponential family if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right).$$

Note: $h(x), t_1(x), \ldots, t_k(x)$ do not depend on $\boldsymbol{\theta}$ and $c(\boldsymbol{\theta})$ does not depend of x.

Example:
Consider $X \sim b(n,p)$ with n fixed. Show that $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$ can be expressed in the exponential family form.

$$\begin{aligned} p(x) &= \binom{n}{x}p^x(1-p)^{n-x} \\ &= \binom{n}{x}\left(\frac{p}{1-p}\right)^x(1-p)^n \\ &= \binom{n}{x}(1-p)^n e^{log(\frac{p}{1-p})^x} \\ &= \binom{n}{x}(1-p)^n e^{xlog(\frac{p}{1-p})} \end{aligned}$$

Therefore this pmf is an exponential family with
$h(x) = \binom{n}{x}, c(p) = (1-p)^n, t_1(x) = x, w_1(p) = log\frac{p}{1-p}$.

Theorem:
Suppose a random variable X has a pdf or pmf that can be expressed in the form of exponential family. Then,

$$\text{(a)} \quad E\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(x)\right) = -\frac{\partial}{\partial \theta_j} logc(\boldsymbol{\theta}).$$

and

$$\text{(b)} \quad var\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(x)\right) = -\frac{\partial^2}{\partial \theta_j^2} logc(\boldsymbol{\theta}) - E\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial \theta_j^2} t_i(x)\right).$$

Note: Here log is the natural logarithm.

Proof of (a):

$$\begin{aligned} \int_x f(x|\boldsymbol{\theta})dx &= 1 \\ \int_x h(x)c(\boldsymbol{\theta})exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) dx &= 1 \end{aligned}$$

Differentiate both sides w.r.t. θ_j:

$$\begin{aligned} &\int_x h(x)\frac{\partial c(\boldsymbol{\theta})}{\partial \theta_j} exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) dx \\ + &\int_x h(x)c(\boldsymbol{\theta}) \sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(x) exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) dx = 0 \end{aligned}$$

Multiply the first integral by $\frac{c(\boldsymbol{\theta})}{c(\boldsymbol{\theta})}$ and note that $\frac{\partial logc(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial c(\boldsymbol{\theta})}{\partial \theta_j}\frac{1}{c(\boldsymbol{\theta})}$.

$$\begin{aligned} &\int_x h(x)\frac{\partial c(\boldsymbol{\theta})}{\partial \theta_j} exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) \frac{c(\boldsymbol{\theta})}{c(\boldsymbol{\theta})} dx \\ + &\int_x h(x)c(\boldsymbol{\theta}) \sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(x) exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right) dx = 0 \end{aligned}$$

After rearranging we get

$$\int_x \sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(x) h(x) c(\boldsymbol{\theta}) exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right) dx = -\frac{\partial logc(\boldsymbol{\theta})}{\partial \theta_j} \int_x h(x) c(\boldsymbol{\theta}) exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right) dx$$

Or

$$E\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(x)\right) = -\frac{\partial}{\partial \theta_j} logc(\boldsymbol{\theta}).$$

To prove statement (b) of the theorem differentiate a second time and rearrange.

Note: $\boldsymbol{\theta}$ can be a single parameter or a vector of parameters.

Show that $X \sim \Gamma(\alpha, \beta)$ and $X \sim$ Poisson(λ) are exponential families.

Use the two statements of the theorem to find the mean and variance of $X \sim \Gamma(\alpha, \beta)$ and $X \sim$ Poisson(λ).

Expectations

- Let X be a discrete random variable. Its expected value (μ), denoted with $E(X)$ is computed using

$$\mu = E(X) = \sum_x xP(x)$$

- Expected value of a function of a discrete random variable: Let $g(X)$ a function of a discrete random variable X. Then its expected value is computed as follows:

$$E[g(X)] = \sum_x g(x)P(x)$$

- Variance (σ^2) of a discrete random variable, denoted with $var(X)$:

$$\begin{aligned} var(X) = \sigma^2 &= E(X-\mu)^2 = \sum_x (x-\mu)^2)P(x) \\ &= \sum_x x^2 P(x) - \mu^2 \end{aligned}$$

The standard deviation of a discrete random variable is the square root of the variance:

$$SD(X) = \sqrt{\sigma^2} = \sqrt{\sum_x (x-\mu)^2)P(x) = \sum_x x^2 P(x) - \mu^2}$$

It follows that:

$$\sigma^2 = EX^2 - \mu^2 \ \text{ or } \ EX^2 = \sigma^2 + \mu^2$$

- Let X be a continuous random variable. Its expected value is equal to

  $$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

- Expected value of a function of a continuous random variable:

  $$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- Variance of continuous random variable:

  $$\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

  or

  $$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - [E(X)]^2$$

- Some properties of expectation and variance (for both discrete and continuous random variables). Let a, b constants and X, Y be random variables.

  $$\begin{aligned} E(X+a) &= a + E(X) \\ E(X+Y) &= E(X) + E(Y) \\ var(X+a) &= var(X) \\ var(aX+b) &= a^2 var(X) \end{aligned}$$

  If X, Y are independent then

  $$var(X+Y) = var(X) + var(Y)$$

The distribution of a function of a random variables

Suppose we know the pdf of a random variable X. Many times we want to find the probability density function (pdf) of a function of the random variable X. Suppose $Y = X^n$.

We begin with the cumulative distribution function of Y:

$$F_Y(y) = P(Y \leq y) = P(X^n \leq y) = P(X \leq y^{\frac{1}{n}}).$$

So far we have

$$F_Y(y) = F_X(y^{\frac{1}{n}})$$

To find the pdf of Y we simply differentiate both sides wrt to y:

$$f_Y(y) = \frac{1}{n} y^{\frac{1}{n}-1} \times f_X(y^{\frac{1}{n}}).$$

where, $f_X(\cdot)$ is the pdf of X which is given. Here are some more examples.

Example 1
Suppose X follows the exponential distribution with $\lambda = 1$. If $Y = \sqrt{X}$ find the pdf of Y.

Example 2
Let $X \sim N(0, 1)$. If $Y = e^X$ find the pdf of Y. Note: Y it is said to have a log-normal distribution.

Example 3
Let X be a continuous random variable with pdf $f(x) = 2(1-x), 0 \leq x \leq 1$. If $Y = 2X - 1$ find the pdf of Y.

Example 4
Let X be a continuous random variable with pdf $f(x) = \frac{3}{2}x^2, -1 \leq x \leq 1$. If $Y = X^2$ find the pdf of Y.

Moment generating functions

Definition:

$$M_X(t) = Ee^{tX}$$

Therefore,
If X is discrete

$$M_X(t) = \sum_x e^{tX} P(x)$$

If X is continuous

$$M_X(t) = \int_x e^{tX} f(x) dx$$

Aside:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Similarly,

$$e^{tx} = 1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots$$

Let X be a discrete random variable.

$$M_X(t) = \sum_x e^{tX} P(x) = \sum_x \left[1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots\right] P(x)$$

or

$$M_X(t) = \sum_x P(x) + \frac{t}{1!} \sum_x x P(x) + \frac{t^2}{2!} \sum_x x^2 P(x) + \frac{t^3}{3!} \sum_x x^3 P(x) + \cdots$$

To find the k_{th} moment simply evaluate the k_{th} derivative of the $M_X(t)$ at $t = 0$.

$$EX^k = [M_X(t)]_{t=0}^{k_{th}\text{derivative}}$$

For example:
First moment:

$$M_X(t)' = \sum_x xP(x) + \frac{2t}{2!}\sum_x x^2P(x) + \cdots$$

We see that $M_X(0)' = \sum_x xP(x) = E(X)$.
Similarly,
Second moment

$$M_X(t)'' = \sum_x x^2P(x) + \frac{6t}{3!}\sum_x x^3P(x) + \cdots$$

We see that $M_X(0)'' = \sum_x x^2P(x) = E(X^2)$.

Examples:
Find the moment generating function of $X \sim b(n, p)$.
Find the moment generating function of $X \sim Poisson(\lambda)$.
Find the moment generating function of $X \sim exp(\lambda)$.
Find the moment generating function of $Z \sim N(0, 1)$.

Theorem:
Let X, Y be independent random variables with moment generating functions $M_X(t), M_Y(t)$ respectively. Then, the moment generating function of the sum of these two random variables is equal to the product of the individual moment generating functions:

$$M_{X+Y}(t) = M_X(t)M_Y(t))$$

Proof:

If X, Y are independent find the distribution of $X + Y$ when.

- $X \sim b(n_1, p), Y \sim b(n_2, p)$
- $X \sim Poisson(\lambda_1), Y \sim Poisson(\lambda_2)$.
- $X \sim N(\mu_1, \sigma_1), Y \sim N(\mu_2, \sigma_2)$.

Properties of moment generating functions:
Let X be a random variable with moment generating function $M_X(t)$. If a, b are constants then

1. $M_{X+a}(t) = e^{at} M_X(t)$
2. $M_{bX}(t) = M_X(bt)$
3. $M_{\frac{X+a}{b}} = e^{\frac{a}{b}t} M_X(\frac{t}{b})$

Proof:

Use these properties and the moment generating function of the standard normal distribution $Z \sim N(0, 1)$ to find the moment generating function of $X \sim N(\mu, \sigma)$.

Let $ln(X) \sim N(\mu, \sigma)$. Find the mean and variance of X. Note: This is called the lognormal distribution.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from $N(\mu, \sigma)$. Use moment generating functions to find the distribution of

a. $T = X_1 + X_2 + \ldots + X_n$.

b. $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.

Joint distributions

- Discrete joint (or bivariate) probability distribution of X and Y: $f_{XY}(x, y) = P(X = x, Y = y)$.
- Always,

$$\sum_x \sum_y f_{XY}(x, y) = 1$$

- Marginal distribution of X:

$$f_X(x) = \sum_y f_{XY}(x, y)$$

- Marginal distribution of Y:

$$f_Y(y) = \sum_x f_{XY}(x, y)$$

- Conditional probability function:

$$f_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)}$$

  Or

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

- Joint cumulative distribution function:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

- Conditional expectation:

$$E[Y|X = x] = \sum_y y f_{Y|X}(y|x)$$

- Conditional variance:

$$var[Y|X = x] = E[Y^2|X = x] - (E[Y|X = x])^2$$

or

$$var[Y|X = x] = \sum_y y^2 f_{Y|X}(y|x) - \left(\sum_y y f_{Y|X}(y|x)\right)^2$$

- Interesting and important property of conditional expectation:

$$E(Y) = E\left[E[Y|X]\right]$$

- Independent discrete random variables:
  X and Y are independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y), \text{ for all pairs of } x, y.$$

- Expected value of a function of X and Y:

$$E\left[g_{X,Y}(x, y)\right] = \sum_x \sum_y g_{X,Y}(x, y) f_{XY}(x, y)$$

- Continuous joint probability distribution of X and Y.
- $f_{XY}(x,y)$ is said to be the joint probability density function of X and Y if

$$\int_y \int_x f_{XY}(x,y)dxdy = 1$$

- Marginal probability density function of X:

$$f_X(x) = \int_y f_{XY}(x,y)dy$$

- Marginal probability density function of Y:

$$f_Y(y) = \int_x f_{XY}(x,y)dx$$

- Conditional probability density function of X given Y:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

- If X, Y are independent, $f_{XY}(x,y) = f_X(x)f_Y(y)$.
- Joint cumulative distribution function (cdf) of (X,Y):

$$\begin{aligned} F_{XY}(x,y) &= P(X \le x, Y \le y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)dtds \end{aligned}$$

  It follows that

$$f_{XY}(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

- Expected value of a function of X and Y:

$$E\left[g_{X,Y}(x,y)\right] = \int_x \int_y g_{X,Y}(x,y) f_{XY}(x,y)dydx$$

Covariance and correlation

Let random variables X, Y with means μ_X, μ_Y respectively. The covariance, denoted with $cov(X, Y)$, is a measure of the association between X and Y.

Definition:

$$\sigma_{XY} = cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

Note: If X, Y are independent then $E(XY) = (EX)E(Y)$ Therefore $cov(X, Y) = 0$. The opposite is NOT always true!

Let W, X, Y, Z be random variables, and a, b, c, d be constants,

- Find $cov(a + X, Y)$
- Find $cov(aX, bY)$
- Find $cov(X, Y + Z)$
- Find $cov(aW + bX, cY + dZ)$

- Important:

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

- Find $var(aX + bY)$

However, the covariance depends on the scale of measurement and so it is not easy to say whether a particular covariance is small or large. The problem is solved by standardize the value of covariance (divide it by $\sigma_X\sigma_Y$), to get the so called coefficient of correlation ρ_{XY}.

$$\rho = \frac{cov(X,Y)}{\sigma_X\sigma_Y}, \quad \text{Always, } -1 \le \rho \le 1,$$

Show that $-1 \le \rho \le 1$:
Let X, Y be random variables with variances σ_X^2, σ_Y^2 respectively. Examine the following random expressions:

$\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$	$\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$

A general result: The covariance operation is additive:

$$\operatorname{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{cov}(X_i, Y_j).$$

Proof:

Let $E(X_i) = \mu_i$ and $E(Y_j) = v_j$. Then

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mu_i \text{ and } E(\sum_{j=1}^{m} Y_j) = \sum_{j=1}^{m} v_j.$$

Therefore,

$$\begin{aligned}
\operatorname{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) &= E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)\left(\sum_{j=1}^{m} Y_j - \sum_{j=1}^{m} v_j\right)\right] \\
&= E\left[\sum_{i=1}^{n}(X_i - \mu_i)\sum_{j=1}^{m}(Y_j - v_j)\right] \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - \mu_i)(Y_j - v_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} E(X_i - \mu_i)(Y_j - v_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \operatorname{cov}(X_i, Y_j).
\end{aligned}$$

Using the result above find $\operatorname{var}(X_1 + \ldots + X_n)$:

$$\begin{aligned}
\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) &= \operatorname{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \operatorname{var}(X_i) + \sum_{i=1}^{n}\sum_{j\neq i}^{n} \operatorname{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \operatorname{var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} \operatorname{cov}(X_i, X_j)
\end{aligned}$$

The previous result can be extended to the more general case:

Let $X_1, X_2, \ldots, X_n$ be random variables, and $a_1, a_2, \ldots, a_n$ be constants. Find the variance of the linear combination $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$.

$$\begin{aligned}
\mathrm{var}\left(\sum_{i=1}^{n} a_i X_i\right) &= \mathrm{cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_j X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} a_i^2 \mathrm{var}(X_i) + \sum_{i=1}^{n}\sum_{j \neq i}^{n} a_i a_j \mathrm{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} a_i^2 \mathrm{var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} a_i a_j \mathrm{cov}(X_i, X_j)
\end{aligned}$$

See also page 24 for the same result expressed in matrix and vector form!

Joint probability distribution of functions of random variables

Let X_1, X_2 be jointly continuous random variables with pdf $f_{X1X2}(x_1, x_2)$. Suppose $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$. We want to find the joint pdf of Y_1, Y_2. We follow this procedure:

1. Solve the equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ for x_1 and x_2 in terms of y_1 and y_2.
2. Compute the Jacobian: $\mathbf{J} = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix}$. ($\mathbf{J}$ is the determinant of the matrix of partial derivatives.)

To find the joint pdf of Y_1, Y_2 use the following result: $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)|\mathbf{J}|^{-1}$, where $|\mathbf{J}|$ is the absolute value of the Jacobian. Here, x_1, x_2 are the expressions obtained from step (1) above.

Example:
Suppose X and Y are independent random variables with $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$. Compute the joint pdf of $U = X + Y$ and $V = \frac{X}{X+Y}$ and find the distribution of U and the distribution of V. Also show that U, V are independent.

Random vectors

Let $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, be a random vector.

Let the mean vector $\boldsymbol{\mu} = E\mathbf{X} = \begin{pmatrix} EX_1 \\ EX_2 \\ \vdots \\ EX_n \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$

and the variance covariance matrix

$$\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \cdots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \cdots & \sigma_{3n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \sigma_{n4} & \cdots & \sigma_n^2 \end{pmatrix}.$$

Theorem:
1. Let $\mathbf{a} = (a_1, \ldots, a_n)'$ be $n \times 1$ vector of constants.
2. Let A be $p \times n$ matrix of constants.

Find the mean and variance of $\mathbf{a}'\mathbf{X}$ and $\mathbf{AX}$.

Joint moment generating functions

Let $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, be a random vector and let $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix}$. The joint moment generating function of $\mathbf{X}$ is defined as $M_{\mathbf{X}}(\mathbf{t}) = Ee^{\mathbf{t}'\mathbf{X}} = Eexp(\sum_{i=1}^{n} t_i x_i)$.

Theorem
Let $M_i(\mathbf{t}) = \frac{\partial M_{\mathbf{X}}(\mathbf{t})}{\partial t_i}$, $M_{ii}(\mathbf{t}) = \frac{\partial^2 M_{\mathbf{X}}(\mathbf{t})}{\partial {t_i}^2}$, and $M_{ij}(\mathbf{t}) = \frac{\partial^2 M_{\mathbf{X}}(\mathbf{t})}{\partial t_i \partial t_j}$.
Then, $EX_i = M_i(\mathbf{0})$, $EX_i^2 = M_{ii}(\mathbf{0})$, and $EX_iX_j = M_{ij}(\mathbf{0})$.

Corollary
Let $\psi(\mathbf{t}) = logM_{\mathbf{X}}(\mathbf{t})$, $\psi_i(\mathbf{t}) = \frac{\partial \psi_{\mathbf{X}}(\mathbf{t})}{\partial t_i}$, $\psi_{ii}(\mathbf{t}) = \frac{\partial^2 \psi_{\mathbf{X}}(\mathbf{t})}{\partial {t_i}^2}$,
and $\psi_{ij}(\mathbf{t}) = \frac{\partial^2 \psi_{\mathbf{X}}(\mathbf{t})}{\partial t_i \partial t_j}$.
Then $EX_i = \psi_i(\mathbf{0})$, $var(X_i) = \psi_{ii}(\mathbf{0})$, and $cov(X_iX_j) = \psi_{ij}(\mathbf{0})$.

Theorem
Let $\mathbf{X} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}$. The marginal moment generating function of $\mathbf{Y}$ ($\mathbf{Z}$) is the moment generating function of $\mathbf{X}$ ignoring the vector $\mathbf{Z}$ ($\mathbf{Y}$). This is expressed as $M_{\mathbf{Y}}(\mathbf{u}) = \mathbf{M}_{\mathbf{X}}(\mathbf{u}, \mathbf{0})$ and $M_{\mathbf{Z}}(\mathbf{v}) = \mathbf{M}_{\mathbf{X}}(\mathbf{0}, \mathbf{v})$, where $\mathbf{t} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$.

Proof

Theorem
If $\mathbf{Y}$ and $\mathbf{Z}$ are independent then $M_{\mathbf{X}}(\mathbf{t}) = \mathrm{M}_{\mathbf{Y}}(\mathbf{u})\mathrm{M}_{\mathbf{Z}}(\mathbf{v})$.
Proof

Example
Let X and Y be independent normal random variables, each with mean μ and standard deviation σ.

a. Consider the random quantities $X + Y$ and $X - Y$. Find the moment generating function of $X + Y$ and the moment generating function of $X - Y$.

c. Find the joint moment generating function of $(X + Y, X - Y)$.

d. Are $X + Y$ and $X - Y$ independent? Explain your answer using moment generating functions.

Multivariate normal distribution

One of the most important distributions in statistical inference is the multivariate normal distribution. The probability density function of the multivariate normal distribution, its moment generating function, and its properties are discussed here.

Probability density function

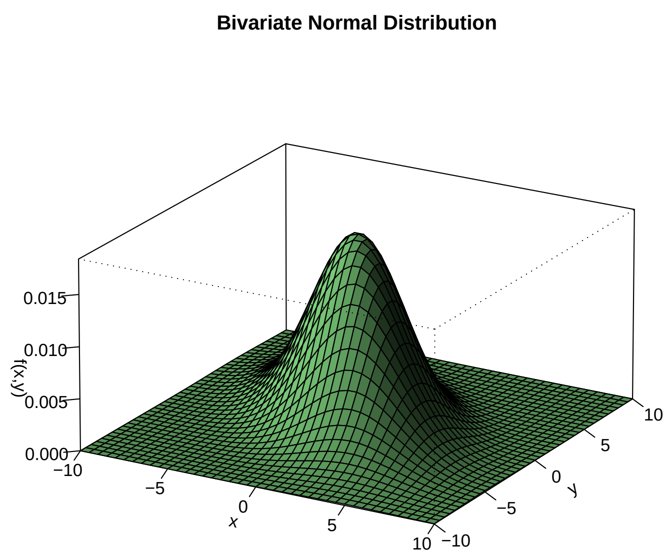
We say that a random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\mathbf{\Sigma}$ follows the multivariate normal distribution if its probability density function is given by

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} |\mathbf{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})},$$

and we write, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$. If $\mathbf{Y} = (Y_1, Y_2)$ then we have a bivariate normal distribution and its probability density function can be expressed as

$$\begin{aligned} f(y_1, y_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \; exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2\mu_2}{\sigma_2}\right)\right]\right] \end{aligned}$$

Here, we have $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$. The previous expression can be obtained by finding the inverse of $\mathbf{\Sigma}$ and substituting it into $f(\mathbf{Y})$.
Here is the bivariate normal pdf.



Moment generating function

A useful tool in statistical theory is the moment generating function. The joint moment generating function is defined as

$$M_{\mathbf{Y}}(\mathbf{t}) = Ee^{\mathbf{t}'\mathbf{Y}} = Ee^{\sum_{i=1}^{n} y_i t_i},$$

where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ and $\mathbf{t} = (t_1, t_2, \ldots, t_n)'$. Suppose $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{I})$. Since $Z_1, Z_2, \ldots, Z_n$ are independent the joint moment generating function of $\mathbf{Z}$ is $M_{\mathbf{Z}}(\mathbf{t}) = e^{\frac{1}{2}\mathbf{t}'\mathbf{t}}$. Why? To find the joint moment generating function of $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ we use the transformation $\mathbf{Y} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$ to get $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$.

Theorem 1

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and let $\mathbf{A}$ be an $m \times n$ matrix of rank m and $\mathbf{c}$ be an $m \times 1$ vector. Then $\mathbf{AY} + \mathbf{c} \sim N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Proof

Theorem 2

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Sub-vectors of $\mathbf{Y}$ follow the multivariate normal distribution and linear combinations of $Y_1, Y_2, \ldots, Y_n$ follow the univariate normal distribution.

Proof

Suppose $\mathbf{Y}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are partitioned as follows $\mathbf{Y} = \begin{pmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, where $\mathbf{Y_1}$ is $p \times 1$. We will show that $\mathbf{Y_1} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{Y_2} \sim N_{n-p}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. The result follows directly by using the previous theorem with $\mathbf{A} = (\mathbf{I_p}, \mathbf{0})$. For a linear combination of $Y_1, Y_2, \ldots, Y_n$, i.e. $a_1Y_1 + a_2Y_2 + \ldots + a_nY_n = \mathbf{a}'\mathbf{Y}$, the matrix $\mathbf{A}$ of theorem 1 is a vector and therefore, $\mathbf{a}'\mathbf{Y} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$.

Example

Let $\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \hline Y_3 \\ Y_4 \\ Y_5 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \hline \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix}, \boldsymbol{\Sigma} = \left(\begin{array}{cc|ccc} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \hline \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_5^2 \end{array}\right)$, then if $\mathbf{Q_1} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$,

it follows that $\mathbf{Q_1} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right]$.

Statistical independence
Suppose $\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ are partitioned as in theorem 2. We say that $\mathbf{Y_1}, \mathbf{Y_2}$ are statistically independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$. This can easily be shown using the joint moment generating function of $\mathbf{Y}$. Recall that the exponent of the joint moment generating function of the multivariate normal distribution is $\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}$ which after partitioning $\mathbf{t}$ conformably (according to the partitioning of $\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$) can be expressed as $\mathbf{t_1}'\boldsymbol{\mu}_1 + \mathbf{t_2}'\boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{t_1}'\boldsymbol{\Sigma}_{11}\mathbf{t_1} + \frac{1}{2}\mathbf{t_2}'\boldsymbol{\Sigma}_{22}\mathbf{t_2} + \mathbf{t_1}'\boldsymbol{\Sigma}_{12}\mathbf{t_2}$. When $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, the joint moment generating function can be expressed as the product of the two marginal moment generating functions of $\mathbf{Y_1}$ and $\mathbf{Y_2}$, i.e. $M_{\mathbf{Y}}(\mathbf{t}) = M_{\mathbf{Y_1}}(\mathbf{t_1})M_{\mathbf{Y_2}}(\mathbf{t_2})$, therefore, $\mathbf{Y_1}$ and $\mathbf{Y_2}$ are independent.

Theorem 3
Using theorem 1 and the statement about statistical independence above, we prove the following theorem. Suppose $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and define the following two vectors $\mathbf{Q_1} = \mathbf{AY}$ and $\mathbf{Q_2} = \mathbf{BY}$. Then, $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are independent if $cov(\mathbf{Q_1}, \mathbf{Q_2}) = \mathbf{A\Sigma B}' = \mathbf{0}$.

Proof
We stack the two vectors as follows: $\mathbf{Q} = \begin{pmatrix} \mathbf{Q_1} \\ \mathbf{Q_2} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{Y} = \mathbf{LY}$. Therefore using theorem 1 we find that $\mathbf{Q} \sim N(\mathbf{L}\boldsymbol{\mu}, \mathbf{L\Sigma L}')$ or

$\mathbf{Q} \sim N\left[\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \boldsymbol{\mu}, \begin{pmatrix} \mathbf{A\Sigma A}' & \mathbf{A\Sigma B}' \\ \mathbf{B\Sigma A}' & \mathbf{B\Sigma B}' \end{pmatrix}\right]$, and we conclude that $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are independent if and only if $\mathbf{A\Sigma B}' = \mathbf{0}$.

Example
Consider the bivariate normal distribution (see page 1). From theorem 1 it follows that $Y_1 \sim N(\mu_1, \sigma_1)$. This is also called the marginal probability distribution of Y_1. We want to find the conditional distribution of Y_2 given Y_1.

From the conditional probability law, $f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1Y_2}(y_1,y_2)}{f_{Y_1}(y_1)}$, and after substituting the bivariate density and the marginal density it can be shown that the conditional probability density function of Y_2 given Y_1 is given by

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{1}{\sqrt{\sigma_2^2(1-\rho)^2}\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{Y_2 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)}{\sigma_2^2(1-\rho^2)}\right)\right].$$

We recognize that this is a normal probability density function with mean
$\mu_{Y_2|Y_1} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(Y_1 - \mu_1)$ and variance $\sigma^2_{Y_2|Y_1} = \sigma_2^2(1-\rho^2)$.

In general:
Suppose that $\mathbf{Y}$, $\boldsymbol{\mu}$, and $\mathbf{\Sigma}$ are partitioned as follows $\mathbf{Y} = \begin{pmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$, $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$, and $\mathbf{Y} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$. It can be shown that the conditional distribution of $\mathbf{Y_1}$ given $\mathbf{Y_2}$ is also multivariate normal, $\mathbf{Y_1}|\mathbf{Y_2} \sim N(\boldsymbol{\mu}_{1|2}, \mathbf{\Sigma}_{1|2})$, where $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{Y_2} - \boldsymbol{\mu}_2)$, and $\mathbf{\Sigma}_{1|2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.

Proof:
Let $\mathbf{U} = \mathbf{Y_1} - \mathbf{\Sigma_{12}}\mathbf{\Sigma_{22}^{-1}}\mathbf{Y_2}$ and $\mathbf{V} = \mathbf{Y_2}$.