Sums of independent random variables and the Convolution formula(s).

We will restrict our attention to the case of two independent continuous r.v.s X and Y.

Suppose X and Y are jointly continuous r.v.s having joint pdf

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y).$$

i.e., the joint pdf factors as the product of its marginals so that X and Y are <u>independent</u>

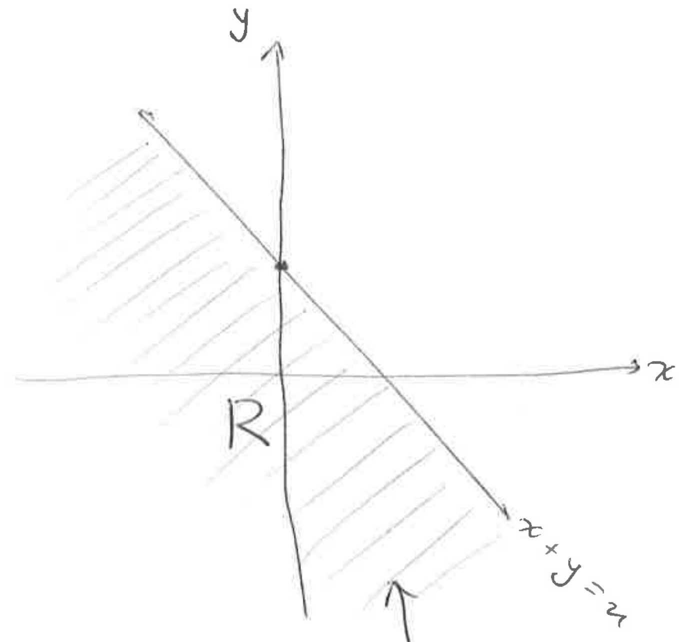
How can we find the pdf of the sum $X+Y$?

Let's use the cdf method.

Define $S = X + Y$.

$$F_S(u) = P(S \le u) = P(X + Y \le u)$$

$$= \iint_R f_{X,Y}(x,y)\, dx\, dy$$



need to integrate the joint pdf of X, Y over this region R

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{u-y} f_X(x) f_Y(y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} f_Y(y) \left\{ \int_{-\infty}^{u-y} f_X(x)\, dx \right\} dy$$

$$= \int_{-\infty}^{\infty} f_Y(y) F_X(u-y)\, dy.$$

So,

$$F_{X+Y}(u) = \int_{-\infty}^{\infty} f_Y(y) F_X(u-y)\, dy.$$

Taking a derivative in u, we find

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_Y(y) \frac{d}{du} F_X(u-y)\, dy = \int_{-\infty}^{\infty} f_X(u-y) f_Y(y)\, dy.$$

i.e. we have the so-called CONVOLUTION formula:

$$\boxed{f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(u-y) f_Y(y)\, dy}$$

<u>Remark</u> By integrating $dy\,dx$ in the above it can be shown that

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(u-x)\, dx$$

So we can use either formula to compute the pdf of $X+Y$:

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(u-x)\, dx$$

$$= \int_{-\infty}^{\infty} f_X(u-y)\, f_Y(y)\, dy$$

and either of these are called the Convolution of f_X and f_Y.

Sometimes written as

$$(f_X * f_Y)(u).$$

To reproduce an example from last lecture ...

Suppose $X \sim \exp\left(\frac{1}{\beta}\right)$ and $Y \sim \underbrace{\exp\left(\frac{1}{\beta}\right)}$ are independent.

Gamma$(1, \beta)$. by our definition.

Use the Convolution formula to find the pdf of their sum $X+Y$.

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(u-x)\, dx$$

$$= \int_0^{\infty} f_X(x)\, f_Y(u-x)\, dx \qquad \text{since } f_X(x) = 0 \text{ if } x < 0$$

$$= \int_0^{u} f_X(x)\, f_Y(u-x)\, dx \qquad \text{since } f_Y(u-x) = 0 \text{ if } x > u.$$

Thus for non-negative r.v.s
the convolution formula can be replaced by

$$f_{X+Y}(u) = \int_0^u f_X(x)\, f_Y(u-x)\, dx.$$

Now $f_X(x) = \frac{1}{\beta} e^{-x/\beta}$ for $x > 0$ (Similarly for y).

Thus,

$$f_{X+Y}(u) = \int_0^u \frac{1}{\beta} e^{-x/\beta} \cdot \frac{1}{\beta} e^{-\frac{(u-x)}{\beta}}\, dx$$

$$= \frac{1}{\beta^2} \int_0^u e^{-u/\beta}\, dx = \frac{u e^{-u/\beta}}{\beta^2} \sim \text{Gamma}(2, \beta).$$

So the sum of two independent $\exp(\frac{1}{\beta})$'s is
a $\text{Gamma}(2, \beta)$.

Covariance of two random variables

Definition

$$\boxed{\mathrm{Cov}(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]}$$

Also, if you multiply out $(X-E(X))(Y-E(Y))$ and use linearity of expected values you get

$$E\left[(X-E(X))(Y-E(Y))\right] = E\left[XY - XE(Y) - E(X)Y + E(X)E(Y)\right]$$

$$= E(XY) - E[XE(Y)] - E(E(X)Y) + E(E(X)E(Y))$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

i.e.,
$$\boxed{\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y).}$$

There are some properties of the Covariance that we will mention.

(1) If $X=Y$, then $\mathrm{Cov}(X,X) = \mathrm{var}(X)$.

(2) $\mathrm{Cov}(X,Y) = \mathrm{cov}(Y,X)$. i.e., Symmetric in its arguments.

(3) $\mathrm{Cov}(a_1 X_1 + a_2 X_2, Y) = a_1 \mathrm{Cov}(X_1, Y) + a_2 \mathrm{Cov}(X_2, Y)$

(4) $\mathrm{Cov}(X, b_1 Y_1 + b_2 Y_2) = b_1 \mathrm{Cov}(X, Y_1) + b_2 \mathrm{Cov}(X, Y_2)$

Properties (3) and (4) are called the multilinearity properties of covariance, and the extend more generally as

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \, \mathrm{cov}(X_i, Y_j)$$

(5) $\mathrm{Cov}(X, c) = 0$ where c is a constant.

Proof. Since $E(c) = c$, we have

$$\mathrm{cov}(X, c) = E(Xc) - E(X)E(c) = cE(X) - cE(X) = 0$$

Remark The covariance between two random variables measures the amount of linear association between them.

To make this notion rigorous I will state the Cauchy - Schwarz inequality.

The Cauchy-Schwarz inequality

for any r.v.s X and Y

$$[E(XY)]^2 \leq E(X^2)\,E(Y^2).$$

or after taking squareroot:

$$|E(XY)| \leq \sqrt{E(X^2)\,E(Y^2)}$$

Proof. For ANY constant c

$$0 \leq E([X - cY]^2) = E(X^2 - 2cXY + c^2Y^2)$$

Or that

$$E(X^2) - 2E(XY)\,c + E(Y^2)\,c^2 \geq 0 \quad (*)$$

which is a polynomial (quadratic) in c. Lets find the c that minimizes this polynomial. Taking derivative in c

$$2E(Y^2)\,c - 2E(XY) = 0$$

$$\Rightarrow \quad c = \frac{E(XY)}{E(Y^2)}.$$

Thus, plugging this back into $(*)$ we have

$$E(X^2) - 2E(XY)\frac{E(XY)}{E(Y^2)} + E(Y^2)\left(\frac{E(XY)}{E(Y^2)}\right)^2 \geq 0$$

multiply through by $E(Y^2)$:

$$E(X^2)E(Y^2) - 2(E(XY))^2 + (E(XY))^2 \geq 0$$

or $$\boxed{(E(XY))^2 \leq E(X^2)E(Y^2).}$$

$\square$

<u>Remark.</u> Replacing X with $X - E(X)$

and Y with $Y - E(Y)$

in the Cauchy-Schwarz inequality gives us

$$(Cov(X,Y))^2 \leq Var(X)Var(Y).$$

or that for any r.v.s.

$$|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}$$

or,

$$\boxed{-1 \leq \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \leq +1}$$

We define the Correlation coefficient $\rho = \rho_{X,Y}$ or $\rho(X,Y)$ as

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}.$$

Example Suppose $Y = aX + b$ with a and b constants so that Y is a (perfect) linear function of X.

Then

$$\rho(X,Y) = \frac{\mathrm{cov}(X, aX+b)}{\sqrt{\mathrm{var}(X)\ \mathrm{var}(aX+b)}} = \frac{a\,\overset{\mathrm{Var}(X)}{\mathrm{Cov}(X,X)}}{\sqrt{\mathrm{var}(X)\cdot a^2\,\mathrm{Var}(X)}}$$

$$= \frac{a}{|a|} = \begin{cases} 1 & \text{if } a > 0 \\ -1 & \text{if } a < 0. \end{cases}$$

In otherwords, when Y is a linear function of X, $\rho(X,Y) = \pm 1$, so that we can think of $\rho(X,Y)$ as measuring the strength of linear relationship between X and Y.

When $\rho(X,Y)=0$ we say the random variables X and Y are uncorrelated.

But since $\rho(X,Y)=0$ iff $\mathrm{cov}(X,Y)=0$ we usually define X, Y uncorrelated to mean:

$$\boxed{E(XY) = E(X)E(Y)}$$

When r.v.s X, Y are independent we learned that they are also uncorrelated:

$E(XY) = E(X)E(Y)$ is a consequence of independence

which implies independent r.v.s are uncorrelated.

But the converse is NOT true:

Consider $X = \begin{cases} +1 & \text{with prob } 1/4 \\ 0 & \text{with prob } 1/2 \\ -1 & \text{with prob } 1/4 \end{cases}$ Then X and X^2 are uncorrelated:

$$E(X \cdot X^2) = E(X^3) = 1^3 \cdot \tfrac{1}{4} + (-1)^3 \cdot \tfrac{1}{4} = 0.$$

and $E(X)=0$ So $E(X \cdot X^2) = E(X)E(X^2)$ ✓

But, $P(X=1, X^2=0) = 0$ while

$P(X=1) = 1/4$ and $P(X^2=0) = \frac{1}{2}$

and $P(X=1, X^2=0) \neq P(X=1)P(X^2=0)$

implies X and X^2 cannot be independent. $\boxtimes$

We also learned when X and Y are independent

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

and more generally

$$(*) \quad Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} var(X_i). \quad \text{when}$$

$X_1, X_2, ..., X_n$ are all independent.

When r.v-s $X_1, X_2, ..., X_n$ are Not independent

then formula $(*)$ above does not hold. and we wish

to investigate:

Start with two r.v-s X, Y that are not necessarily

independent.

$$\mathrm{Var}(X+Y) = \mathrm{Cov}(X+Y, X+Y)$$

$$= \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X) + \mathrm{Cov}(Y,Y)$$

$$\boxed{\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y).}$$

since $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$

In general, for r.v.s $X_1, X_2, \ldots, X_n$ we have

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{\substack{i=1 \\ i=j}}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

or

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

$\boxtimes$

Moment-generating functions (mgf)

Let X be a r.v.

The function of s:

$$M(s) := E(e^{sX}) \quad \text{is called the mgf of } X.$$

when this expectation exists and is finite in an open interval containing $s=0$.

Remark 1

The reason this is called the Moment-generating function is the following:

Take one derivative in s

$$M'(s) = E(X e^{sX})$$

taking two derivatives in s:

$$M''(s) = E(X^2 e^{sX})$$

and in general, after n derivatives in s:

$$M^{(n)}(s) = E(X^n e^{sX})$$

Then Substituting $s=0$ into these derivatives we have

$$M^{(n)}(0) = E(X^n) \quad \text{for } n=1,2,3,\ldots$$

i.e., we can recover all the Moment of the r.v. X via the mgf.

Remark 2

The importance of the mgf of X is the following

When two r.v.s X and Y have the same mgf, then their probability distributions are the same!

That is, the <u>mgfs</u> <u>uniquely</u> <u>identify</u> <u>a</u> <u>probability</u> <u>distribution</u>.

Remark 3

Since

$$e^{sX} = 1 + sX + \frac{s^2}{2}X^2 + \frac{s^3}{3!}X^3 + \frac{s^4}{4!}X^4 + \cdots$$

we have

$$M(s) = E\left(e^{sX}\right) = 1 + sE(X) + \frac{s^2}{2}E(X^2) + \frac{s^3}{3!}E(X^3) + \frac{s^4}{4!}E(X^4) + \cdots$$

The Mgf (when it exists) has all the information about it distribution encoded in it!

Let's now compute some mgfs of some of the named distributions we have been working with...

The Bernoulli(p)

$$P(X=0) = 1-p, \quad P(X=1) = p.$$

then

$$\underbrace{E(e^{sX})}_{} = e^{s\cdot 0} P(X=0) + e^{s\cdot 1} P(X=1)$$

$$\boxed{M(s) = 1-p + pe^s}$$

The exponential (λ).

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{for } x > 0.$$

$$M(s) = E(e^{sX}) = \int_0^\infty e^{sx} \lambda e^{-\lambda x}\, dx$$

$$= \int_0^\infty \lambda e^{-(\lambda - s)x}\, dx = \frac{-\lambda}{\lambda - s} e^{-(\lambda-s)x}\Big]_{x=0}^{x=\infty}$$

$$= \frac{\lambda}{\lambda - s} = \left(\frac{\lambda - s}{\lambda}\right)^{-1} = \boxed{\left(1 - \frac{s}{\lambda}\right)^{-1} = M(s)}$$

The geometric(p).

$$P(X=x) = p(1-p)^{x-1} \quad \text{for } x=1,2,3,\ldots$$

then

$$E(e^{sX}) = \sum_{x=1}^{\infty} e^{sx} p(1-p)^{x-1}$$

$$= \frac{p}{1-p}\left\{ \sum_{x=1}^{\infty} \{e^s(1-p)\}^x \right\}$$

$$= \cancel{\frac{p}{1-p}}\left\{ \frac{e^s \cancel{(1-p)}}{1-e^s(1-p)} \right\} = \boxed{\frac{pe^s}{1-(1-p)e^s} = M(s)}$$

For some of these named distributions the mgf can be used to make some powerful connections to other distributions.

Recall the when r.vs X, Y are independent, then

$$E(g(X)h(Y)) = E(g(X))\,E(h(Y)) \quad \text{for any functions}$$

g and h. In particular, when X, Y independent

$$\underbrace{E(e^{s(X+Y)})}_{M_{X+Y}(s)} = E(e^{sX}e^{sY}) = \underbrace{E(e^{sX})}_{M_X(s)}\underbrace{E(e^{sY})}_{M_Y(s)}.$$

and, in general, if $X_1, X_2, \dots, X_n$ are independent then

$$M_{X_1+X_2+\dots+X_n}(s) = M_{X_1}(s) \cdot M_{X_2}(s) \cdots M_{X_n}(s)$$

i.e.,

$$M_{\sum_{i=1}^{n} X_i}(s) = \prod_{i=1}^{n} M_{X_i}(s).$$

Let's now compute the mgf of a binomial(n, p)

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, \dots, n.$$

$$M(s) = E(e^{sX}) = \sum_{x=0}^{n} e^{sx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (e^s p)^x (1-p)^{n-x}$$

$$\boxed{M_X(s) = (1-p+pe^s)^n}$$

Continuing....

Suppose $X_1, X_2, \ldots, X_n$ are independent Bernoulli (p) r.v.s.

Let $Y = X_1 + X_2 + \cdots + X_n$.

Then

$$M_Y(s) = M_{X_1}(s) M_{X_2}(s) \cdots M_{X_n}(s)$$

$$= \underbrace{(1-p+pe^s)\cdot(1-p+pe^s)\cdots(1-p+pe^s)}_{n \text{ times}}$$

$$= (1-p+pe^s)^n.$$

Note: $M_Y(s) = M_X(s)$ where $X \sim \text{binomial}(n, p)$.

By Remark 2 we must have that X and Y have the same probability distribution!

That is, a binomial(n, p) can be thought of as the sum of n independent Bernoulli(p) rvs!