As we've seen joint cdfs, pdfs, pmfs can be quite complicated.

To get insight, might be useful to have summary statistics such as moments like mean, variance, etc. we had for marginal cdf/pdf/etc.

Define moment quantities for X and Y (with finite second moments themselves).

<u>correlation</u> $E[XY]$

<u>covariance</u> $\text{cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$

<u>correlation coefficient</u>: $\rho_{X,Y} = \dfrac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$.

note that $\text{cov}(X,X) = \text{var}(X)$.

just like $\text{var}(X) = E[X^2] - E[X]^2$

we get
$$\text{cov}(X,Y) = E[XY] - E[X]E[Y].$$

if either X or Y is zero-mean, $E[XY] = \text{cov}(X,Y)$.

Random variables X, Y <u>uncorrelated</u> if $\text{cov}(X,Y) = 0$.

equivalently (assuming $\text{var}(X), \text{var}(Y) > 0$,) $\rho_{XY} = 0$.

If $\text{cov}(X,Y) > 0$ or $\rho_{XY} > 0$, then positively correlated

< 0 negatively correlated.

[illegible] joint cdfs, pdfs (pmfs) can be quite complicated. [illegible] summary [illegible]

Define mixed moments for X and Y [illegible] moments [illegible].

<u>Correlation</u>: $E[XY]$

<u>Covariance</u>: $\mathrm{cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$

<u>Correlation coefficient</u>: $\rho_{XY} = \dfrac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$.

Note that $\mathrm{cov}(X,X) = \mathrm{Var}(X)$.

Just like $\mathrm{var}(X) = E[X^2] - E[X]^2$

we get

$\mathrm{cov}(X,Y) = E[XY] - E[X]E[Y]$.

If either X or Y is zero-mean, $E[XY] = \mathrm{cov}(X,Y)$

Random variables X, Y <u>uncorrelated</u> if $\mathrm{cov}(X,Y) = 0$.

equivalently (assuming $\mathrm{var}(X), \mathrm{var}(Y) \neq$) $\rho_{XY} = 0$.

If $\mathrm{cov}(X,Y) > 0$ or $\rho_{XY} > 0$, then positively correlated

< 0 negatively correlated.

If $X \perp\!\!\!\perp Y$, then $E[XY] = E[X]E[Y]$ so uncorrelated

but uncorrelated $\not\Rightarrow$ independence.

Three or more r.v.s uncorrelated if pairwise uncorrelated

(note different than mutual independence where more conditions needed)

Linearity

$cov(X+Y, U+V) = cov(X,U) + cov(X,V) + cov(Y,U) + cov(Y,V)$.

$cov(aX+b, cY+d) = ac\; cov(X,Y)$. for constants a,b,c,d.

Variance of uncorrelated variables

not just for independence, but if X, Y uncorrelated:

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

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

ex Simplify expression

(a) $cov(8X+3, 5Y-2) = cov(8X, 5Y) = 40\, cov(X,Y)$.

(d) $\rho_{10X, Y+4}$: $cov(10X, Y+4) = 10\, cov(X,Y)$.

Further, standard deviation of $10X$ is $10\sigma_X$ and of $Y+4$ is σ_Y

so $\rho_{10X,Y+4} = \dfrac{10\, cov(X,Y)}{(10\sigma_X)(\sigma_Y)} = \dfrac{cov(X,Y)}{\sigma_X \sigma_Y} = \rho_{XY}$.

If $X \perp Y$, then $E[XY] = E[X]E[Y]$ is equivalent to uncorrelated $\not\Rightarrow$ independence.

There are some cases uncorrelated $\Rightarrow$ independent (like when X, Y are jointly normal ...)

Properties

$\operatorname{cov}(X+Y, U+V) = \operatorname{cov}(X,U) + \operatorname{cov}(X,V) + \operatorname{cov}(Y,U) + \operatorname{cov}(Y,V)$

$\operatorname{cov}(aX, bY) = ab\,\operatorname{cov}(X,Y)$

Variance of sum

if X, Y independent ...

$\operatorname{var}(X+Y) = \operatorname{cov}(X+Y, X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y)$

$= \operatorname{var}(X) + \operatorname{var}(Y)$

Useful properties

(a) $\operatorname{cov}(aX+b, cY+d) = ac\,\operatorname{cov}(X,Y)$

(b) $\operatorname{corr}(aX+b, cY+d) = \operatorname{corr}(X,Y)$

(since standard dev of $aX+b$ is $|a|\sigma_X$ ...)

$$\operatorname{corr}(aX+b, cY+d) = \frac{ac\,\operatorname{cov}(X,Y)}{|a|\,|c|\,\sigma_X\sigma_Y} = \operatorname{corr}(X,Y)$$

Generally note that ρ_{XY} is scaled version of $\text{cov}(X,Y)$

It is dimensionless.

In particular

$$\rho_{aX+b,\,cY+d} = \rho_{X,Y} \quad \text{for } a,c>0.$$

ρ is always in interval $[-1,1]$.

To prove, we use Schwarz Inequality:

Thm: for two random variables X and Y,

$$|E[XY]| \leq \sqrt{E[X^2]\,E[Y^2]}\,.$$

Further, if $E[X^2]\neq 0$, equality holds iff $\Pr[Y=cX]=1$ for some constant c.

proof: Let $\lambda = \dfrac{E[XY]}{E[X^2]}$ and note that

$$0 \leq E[(Y-\lambda X)^2] = E[Y^2] - 2\lambda E[XY] + \lambda^2 E[X^2]$$
$$= E[Y^2] - \frac{E[XY]^2}{E[X^2]}$$

which implies $E[XY]^2 \leq E[X^2]\,E[Y^2]$ which is equivalent to Schwarz Ineq.

If $P(Y=cX)=1$ for some c, equality holds with $E[(Y-\lambda X)^2]=0$.

Generally note that ρ_{XY} is scale invariant unlike $cov(X,Y)$

i.e. dimensionless.

In particular

$$\rho_{aX+b,\,cY+d} = \rho_{XY} \quad \text{for } a,c>0.$$

ρ is always in interval $[-1,1]$.

To prove, we use Schwarz inequality

Thm: For two random variables X and Y

$$|E[XY]| \le \sqrt{E[X^2]\,E[Y^2]}$$

Further, if $E[X^2]\neq 0$, equality holds iff $P[Y=cX]=1$ for some c.

Proof: Let $\lambda = \dfrac{E[XY]}{E[X^2]}$ and note that

$$0 \le E[(Y-\lambda X)^2] = E[Y^2] - 2\lambda E[XY] + \lambda^2 E[X^2]$$

$$= E[Y^2] - \frac{E[XY]^2}{E[X^2]}$$

which implies $E[XY]^2 \le E[X^2]\,E[Y^2]$ which is equivalent to Schwarz ineq.

If $P[Y=\lambda X]=1$ for some λ, equality holds since $E[(Y-\lambda X)^2]=0$.

<u>corollary</u>: for two r.v. X and Y:

$$|\mathrm{cov}(X,Y)| \le \sqrt{\mathrm{var}(X)\mathrm{var}(Y)}$$

with equality iff $\mathrm{var}(X) \neq 0$ and $Y = aX + b$ for some constants a, b.

(follows from Schwarz ineq).

Suppose $X_1, \dots, X_n$ are iid with mean μ and variance σ^2.

These are unknown and even the distribution unknown, so can't do MLE.

estimate using sample mean and sample variance:

$$\hat{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \hat{X})^2.$$

(a) Is sample mean unbiased?

$$E[\hat{X}] = \frac{1}{n}\sum_{k=1}^{n} E[X_k] = \frac{1}{n}\sum_{k=1}^{n} \mu = \mu \qquad \text{so yes.}$$

(b) what is MSE for estimating mean by sample mean? $E[(\mu - \hat{X})^2]$.

$$E[(\mu-\hat{X})^2] = \mathrm{var}(\hat{X}) = \frac{1}{n^2}\mathrm{var}\left(\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{var}(X_k) = \frac{1}{n^2}\sum_{k=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.$$

(c) is sample variance unbiased?

$$E[\hat{\sigma}^2] = \frac{1}{n-1}\sum_{k=1}^{n} E[(X_k - \hat{X})^2] = \frac{n}{n-1}E[(X_1 - \hat{X})^2] \quad \text{by symmetry.}$$

now $E[X_1 - \hat{X}] = \mu - \mu = 0$ so

$$E[(X_1-\hat{X})^2] = \mathrm{var}(X_1 - \hat{X}) = \mathrm{var}\left(\frac{(n-1)}{n}X_1 - \sum_{k=2}^{n}\frac{X_k}{n}\right)$$

$$= \left(\left(\frac{n-1}{n}\right)^2 + \sum_{k=2}^{n}\frac{1}{n^2}\right)\sigma^2 = \frac{(n-1)}{n}\sigma^2$$, so $E[\hat{\sigma}^2] = \frac{n}{n-1}\cdot\frac{(n-1)}{n}\cdot\sigma^2 = \sigma^2$ so unbiased