Joint Probability Distributions

[Navidi sections 2.5 and 2.6; Devore sections 5.1-5.2]

The **joint probability mass function** of two discrete random quantities $X,\ Y$ is

$$p(x, y) = \mathrm{P}[X = x \ \textbf{and} \ Y = y]$$

The **marginal probability mass functions** are

$$\boxed{p_X(x) = \sum_y p(x, y)} \qquad \boxed{p_Y(y) = \sum_x p(x, y)}$$

Example 7.01

Find the marginal *p.m.f.*s for the following joint *p.m.f.*

$p(x,y)$	$y = 3$	$y = 4$	$y = 5$	$p_X(x)$
$x = 0$	.30	.10	.20	
$x = 1$	.20	.05	.15	
$p_Y(y)$				

Like any other probability mass function,
a joint p.m.f. must be non-negative for all (x, y) and be coherent, $\displaystyle\sum_{\forall x}\sum_{\forall y} p(x, y) = 1$.

It is easy to check that both conditions are satisfied in example 7.01.

The random quantities X and Y are **independent** if and only if

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad \forall (x, y)$$

An equivalent statement is $\mathrm{P}[(X = x) \cap (Y = y)] = \mathrm{P}[X = x] \cdot \mathrm{P}[Y = y] \quad \forall (x, y)$

In example 7.01, $p_X(0) \cdot p_Y(4) = .60 \times .15 = .09$, but $p(0,4) = .10$.
Therefore X and Y are **dependent**,
[despite $p(x,3) = p_X(x) \cdot p_Y(3)$ for $x = 0$ and $x = 1$!].

For any two possible events A and B, conditional probability is defined by

$\mathrm{P}[A \mid B] = \dfrac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]}$, which leads to the conditional probability mass functions

$$p_{Y|X}(y \mid x) = \frac{p(x,y)}{p_X(x)} \quad \text{and} \quad p_{X|Y}(x \mid y) = \frac{p(x,y)}{p_Y(y)} .$$

In example 7.01,

$p(x,y)$	$y = 3$	$y = 4$	$\boldsymbol{y = 5}$	$p_X(x)$
$\boldsymbol{x = 0}$	.30	.10	**.20**	**.60**
$x = 1$	.20	.05	.15	.40
$p_Y(y)$	.50	.15	.35	1

Joint Probability Density Functions [for bonus questions only]

The continuous analogue to the discrete joint probability mass function is the joint probability density function, $f(x, y)$.

It must satisfy the conditions $f(x,y) \geq 0 \quad \forall (x,y)$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1$.

Joint density functions are related to probability statements by integration over intervals:

$$\text{P}\left[(a < X < b) \cap (c < Y < d)\right] = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

Marginal p.d.f.s are defined by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Conditional p.d.f.s are defined by

$$f_{Y|X}(y\,|\,x) = \frac{f(x,y)}{f_X(x)} \quad \text{and} \quad f_{X|Y}(x\,|\,y) = \frac{f(x,y)}{f_Y(y)}.$$

Two continuous distributions are independent if and only if

$$f(x,y) = f_X(x)\cdot f_Y(y) \quad \forall (x,y)$$

Example 7.02 (Navidi, exercises 2.6, page 156, question 20) [bonus question only]

Let X denote the amount of shrinkage (in %) undergone by a randomly chosen fibre of a certain type when heated to a temperature of 120°C. Let Y represent the additional shrinkage (in %) when the fibre is heated to 140°C. The joint probability density function of X and Y is given by

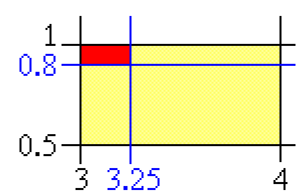
$$f(x,y) = \begin{cases} \dfrac{48xy^2}{49} & (3 < x < 4 \text{ and } 0.5 < y < 1) \\ 0 & (\text{otherwise}) \end{cases}$$

(a) Find $\text{P}[X < 3.25 \text{ and } Y > 0.8]$.

(b) Find the marginal probability density functions $f_X(x)$ and $f_Y(y)$.

(c) Are X and Y independent? Explain.

(a) $\text{P}[X < 3.25 \text{ and } Y > 0.8] = \int_{0.8}^{\infty}\int_{-\infty}^{3.25} f(x,y)\,dx\,dy$



$$= \int_{0.8}^{1}\int_{3}^{3.25} \frac{48xy^2}{49}\,dx\,dy = \frac{48}{49}\int_3^{3.25} x\,dx \cdot \int_{0.8}^{1} y^2\,dy$$

$$= \frac{48}{49}\left[\frac{x^2}{2}\right]_3^{3.25}\left[\frac{y^3}{3}\right]_{0.8}^{1} = \frac{48}{49}\left(\frac{10.5625-9}{2}\right)\left(\frac{1-0.512}{3}\right) = \frac{61}{490} \approx \underline{\underline{\mathbf{.124}}}$$

Example 7.02 (continued)

(b)

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{0.5}^{1} \frac{48xy^2}{49}\,dy = \frac{48}{49}x\left[\frac{y^3}{3}\right]_{0.5}^{1} = \frac{16}{49}x\left(1-\frac{1}{8}\right) = \frac{2x}{7}$$

(for $3 < x < 4$ only) and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{3}^{4} \frac{48xy^2}{49}\,dx = \frac{48}{49}y^2\left[\frac{x^2}{2}\right]_{3}^{4} = \frac{24}{49}y^2(16-9) = \frac{24y^2}{7}$$

(for $0.5 < y < 1$ only).

(c) For all (x, y) such that $3 < x < 4$ and $0.5 < y < 1$,

$$f_X(x)\cdot f_Y(y) = \frac{2x}{7}\cdot\frac{24y^2}{7} = \frac{48xy^2}{49} = f(x,y)$$

Therefore **yes**, X and Y are independent.

Whenever $f(x,y) = g(x)\cdot h(y)\ \forall (x,y)$ on a rectangular domain aligned with the coordinate axes (or on all of $\mathbb{R}^2$), the random quantities X and Y are independent.

These concepts of joint probability distributions (both discrete and continuous) can be extended to the cases of three or more random quantities.

Expected Value

$$\mathrm{E}[h(X,Y)] = \sum_x \sum_y h(x,y) \cdot p(x,y) \quad \text{or} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\, f(x,y)\, dx\, dy$$

A measure of linear dependence is the **covariance** of X and Y:

$$\boxed{\mathrm{Cov}[X,Y] = \mathrm{E}\big[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\big] = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)\, p(x,y)}$$

or, in the continuous case,

$$\mathrm{Cov}[X,Y] = \mathrm{E}\big[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\big] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f(x,y)\, dx\, dy$$

Manipulate the double summation:

$$\mathrm{Cov}[X,Y] = \sum_x \sum_y (xy - \mu_X y - x\mu_Y + \mu_X \mu_Y) p(x,y)$$

$$= \sum_x \sum_y xy \cdot p(x,y) - \mu_X \sum_y y \cdot \sum_x p(x,y) - \mu_Y \sum_x x \cdot \sum_y p(x,y) + \mu_X \mu_Y \sum_x \sum_y p(x,y)$$

$$= \mathrm{E}[XY] - \mu_X \mathrm{E}[Y] - \mu_Y \mathrm{E}[X] + \mu_X \mu_Y$$

Therefore

$$\boxed{\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]}$$

for both discrete and continuous random quantities.

Note that $\mathrm{V}[X] = \mathrm{Cov}[X, X]$.

In <u>Example 7.01</u>, find the covariance of X and Y:

$p(x,y)$	$y = 3$	$y = 4$	$y = 5$	$p_X(x)$
$x = 0$	.30	.10	.20	.60
$x = 1$	.20	.05	.15	.40
$p_Y(y)$	.50	.15	.35	1

Example 7.01 (continued)

$\mathrm{E}[X] =$

$\mathrm{E}[Y] =$

$\mathrm{E}[XY] =$

$x \times y$	**3**	**4**	**5**
0	$0 \times 3 \times .30$	$0 \times 4 \times .10$	$0 \times 5 \times .20$
1	$1 \times 3 \times .20$	$1 \times 4 \times .05$	$1 \times 5 \times .15$

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]$$

Note that the covariance depends on the units of measurement. If X is re-scaled by a factor c and Y by a factor k, then

$$\mathrm{Cov}[cX, kY] = \mathrm{E}[cX\,kY] - \mathrm{E}[cX] \cdot \mathrm{E}[kY] = ck\,\mathrm{E}[XY] - c\,\mathrm{E}[X] \cdot k\,\mathrm{E}[Y]$$

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

A special case is $\mathrm{V}[cX] = \mathrm{Cov}[cX, cX] = c^2 \mathrm{V}[X]$.

This dependence on the units of measurement of the random quantities can be eliminated by dividing the covariance by the product of the standard deviations of the two random quantities.

The **correlation coefficient** of X and Y is $\operatorname{Corr}(X,Y) = \rho_{X,Y} =$

$$\boxed{\rho = \frac{\operatorname{Cov}[X,Y]}{\sqrt{\mathrm{V}[X] \bullet \mathrm{V}[Y]}} = \frac{\mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]}{\sigma_X \cdot \sigma_Y}}$$

In Example 7.01,

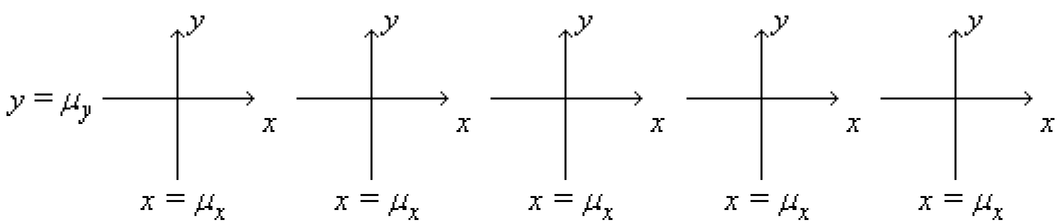
$\mathrm{E}\left[X^2\right] =$

$\mathrm{V}[X] =$

$\mathrm{E}\left[Y^2\right] =$

$\mathrm{V}[Y] =$

<u>Example 7.02</u> For a joint uniform probability distribution:
(and noting $\rho \propto \sum_x \sum_y (x-\mu_x)(y-\mu_y)$):



In general, for constants *a*, *b*, *c*, *d*, with a and c both positive or both negative,

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

Also: $-1 \le \rho \le +1.$

Rule of thumb:

$$|\rho| \geq .8 \quad \Rightarrow \quad \text{strong correlation}$$

$$.5 < |\rho|. < 8 \quad \Rightarrow \quad \text{moderate correlation}$$

$$|\rho| \leq .5 \quad \Rightarrow \quad \text{weak correlation}$$

In the example above, $\rho = 0.0224 \Rightarrow$ very weak correlation (almost uncorrelated).

X, Y are independent $\Rightarrow$ $p(x, y) =$

$$\Rightarrow \quad \mathrm{E}[XY] = \sum\sum xy \cdot p(x) \cdot p(y) = \sum x \cdot p(x) \cdot \sum y \cdot p(y) = \mathrm{E}[X] \cdot \mathrm{E}[Y]$$

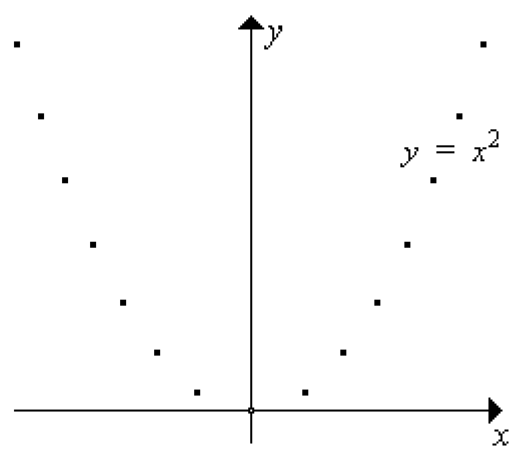
$\Rightarrow$ $\mathrm{Cov}[X, Y] =$

It then follows that
X, Y are independent $\Rightarrow$ X, Y are uncorrelated ($\rho = 0$), but

X, Y are uncorrelated $\not\Rightarrow$ X, Y are independent .

Counterexample (7.03):

Let the points shown be equally likely. Then the value of Y is completely determined by the value of X. The two random quantities are thus highly dependent. Yet they are uncorrelated!



Linear Combinations of Random Quantities

Let the random quantity Y be a linear combination of n random quantities X_i :

$$Y = \sum_{i=1}^{n} a_i X_i$$

then $\mathrm{E}[Y] = \mathrm{E}\left[\sum_{i=1}^{n} a_i X_i\right] =$

But it can be shown that $\boxed{\mathrm{V}[Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{Cov}[X_i, X_j]}$

$\{X_i\}$ independent $\Rightarrow \boxed{\mathrm{V}[Y] = \sum_{i=1}^{n} a_i^2 \mathrm{V}[X_i]}$

Special case: $n = 2,\ a_1 = 1,\ a_2 = \pm 1:$

$\mathrm{E}[X_1 \pm X_2] =$ and

$\mathrm{V}[X_1 \pm X_2] =$

<u>Example 7.04</u>

Two runners' times in a race are independent random quantities T_1 and T_2 , with

$\mu_1 = 40$ $\sigma_1^2 = 4$,

$\mu_2 = 42$ $\sigma_2^2 = 4$,

Find $\mathrm{E}[T_1 - T_2]$ and $\mathrm{V}[T_1 - T_2]$.

Distribution of the Sample Mean

If a random sample of size n is taken and the observed values are $\{ X_1, X_2, X_3, \ldots, X_n \}$ then the X_i are **independent and identically distributed** (***iid***) (each with population mean μ and population variance σ^2) and two more random quantities can be defined:

Sample total:

$$T = \sum_i X_i$$

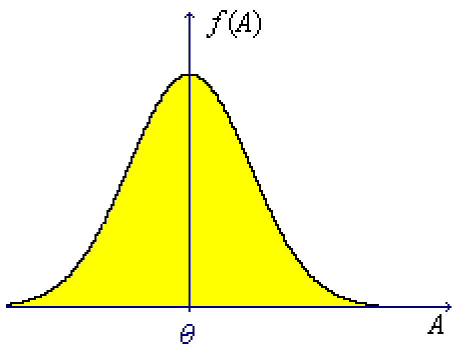
Sample mean:

$$\overline{X} = \frac{T}{n} = \frac{1}{n}\sum_i X_i$$

$\mathrm{E}\left[\overline{X}\right] =$

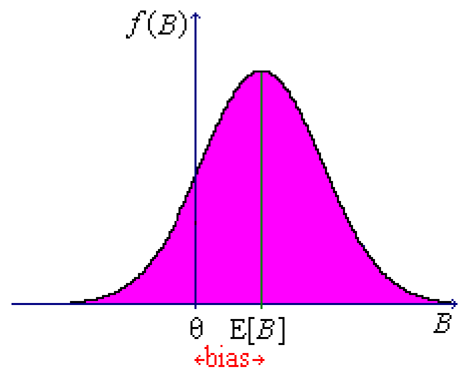
Also $\mathrm{V}\left[\overline{X}\right] =$

Unbiased estimator A for some unknown parameter θ:



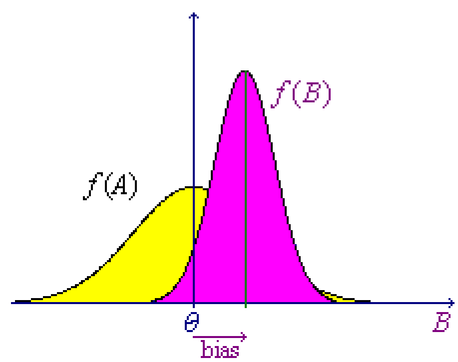
$$\text{E}[A] = \theta$$

Biased estimator B for the unknown parameter θ:



$$\text{E}[B] \neq \theta$$

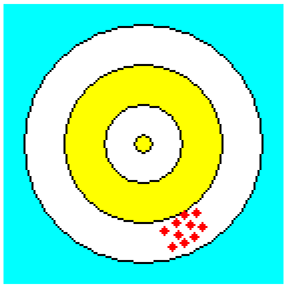
Which estimator should we choose to estimate θ?

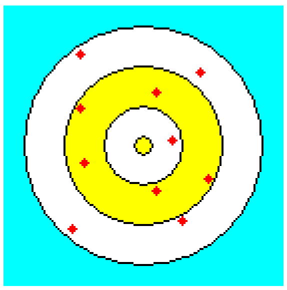


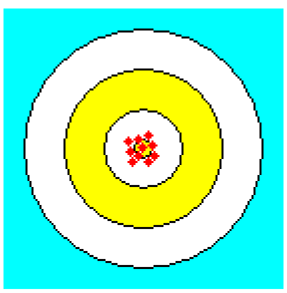
A minimum variance unbiased estimator is ideal.
[See also Problem Set 7 Question 8]

Accuracy and Precision (Example 7.05) [Navidi section 3.1; Devore section 6.1]

An archer fires several arrows at the same target.







Error = Systematic error + Random error

$(\text{error})^2 = (\text{bias})^2 + \text{V}[\text{estimator}]$

Estimator A for θ is consistent iff

$$\text{E}[A] \to \theta \quad \text{and} \quad \text{V}[A] \to 0$$

(as $n \to \infty$)

A particular value a of an **estimator** A is an **estimate**.

Sample Mean

A random sample of *n* values $\{ X_1, X_2, X_3, \ldots, X_n \}$ is drawn from a population of mean μ and standard deviation σ.

Then $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$ and the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$\bar{X}$ estimates μ.

$$\Rightarrow \quad \mathrm{E}[\bar{X}] = \mu, \qquad \mathrm{V}[\bar{X}] = \frac{\sigma^2}{n}$$

But, if μ is unknown, then σ^2 is unknown (usually).

Sample Variance

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1}$$

$$= \frac{n\sum X_i^2 - \left(\sum X_i\right)^2}{n(n-1)}$$

and the **sample standard deviation** is $S = \sqrt{S^2}$

$n - 1$ = number of **degrees of freedom** for S^2.

Justification for the divisor ($n - 1$) [not examinable]:

Using $\mathrm{V}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2$ for all random quantities Y,

$\mathrm{E}[Y^2] = \mathrm{V}[Y] + (\mathrm{E}[Y])^2 = \sigma_Y^2 + \mu_Y^2$

$$\Rightarrow \quad \mathrm{E}\left[\sum_i (X_i - \bar{X})^2\right] = \mathrm{E}\left[\left(\sum_i X_i^2\right) - \frac{1}{n}\left(\sum_i X_i\right)^2\right]$$

$$= \mathrm{E}\left[\sum_i X_i^2\right] - \mathrm{E}\left[\frac{1}{n}\left(\sum_i X_i\right)^2\right] = \underbrace{\sum_i \mathrm{E}\left[X_i^2\right]}_{\text{set } Y = X_i} - \underbrace{\frac{1}{n}\mathrm{E}\left[\left(\sum_i X_i\right)^2\right]}_{\text{set } Y = \sum X_i}$$

$$= \left\{\sum_i \left(\mathrm{V}[X_i] + (\mathrm{E}[X_i])^2\right)\right\} - \frac{1}{n}\left\{\mathrm{V}\left[\sum_i X_i\right] + \left(\mathrm{E}\left[\sum_i X_i\right]\right)^2\right\}$$

$$= \left\{\sum \left(\sigma^2 + \mu^2\right)\right\} - \frac{1}{n}\left\{n\sigma^2 + (n\mu)^2\right\} \qquad (\because \ i.i.d.)$$

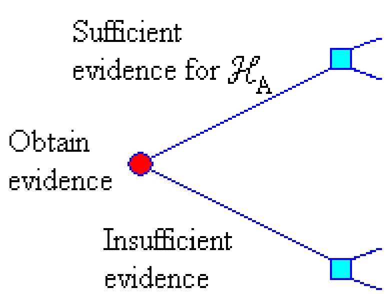
$$= n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2} = (n-1)\sigma^2$$

S^2 is the minimum variance unbiased estimator of σ^2 and
$\overline{X}$ is the minimum variance unbiased estimator of μ.
Both estimators are also consistent.

Inference – Some Initial Considerations

Is a **null hypothesis** $\mathcal{H}_o$ true (our "default belief"), or do we have sufficient evidence to reject $\mathcal{H}_o$ in favour of the **alternative hypothesis** $\mathcal{H}_A$?

$\mathcal{H}_o$ could be "defendant is not guilty" or "$\mu = \mu_o$", etc.
The corresponding $\mathcal{H}_A$ could be "defendant is guilty" or "$\mu \neq \mu_o$", etc.
The **burden of proof is on** $\mathcal{H}_A$ **.**



Bayesian analysis: μ is treated as a random quantity. Data are used to modify prior belief about μ. Conclusions are drawn using both old and new information.
Classical analysis: Data are used to draw conclusions about μ, without using any prior information.

[End of Chapter 7]