5 - Random vectors

- The joint probability matrix
- Marginal random variables
- Independent random variables
- The conditional pmf
- Example: four dice
- Example: dependent events
- Covariance and correlation
- Affine transformations of random vectors
- Sum of two independent random variables
- IID random variables
- Law of large numbers
- Estimating the mean
- Identifying a pmf from data

Random vectors

A *random vector* $x : \Omega \to \mathbb{R}^n$ is a random variable whose values are vectors.

$$x(\omega) = \begin{bmatrix} x_1(\omega) \\ x_2(\omega) \\ \vdots \\ x_n(\omega) \end{bmatrix}$$

- Equivalently, we have n real-valued random variables $x_1, \dots, x_n$, called *marginal random variables*
- The induced pmf $p^x : \mathbb{R}^n \to [0, 1]$ is defined by

$$p^x(a) = \mathbf{Prob}\Big(\{\, \omega \in \Omega \mid x(\omega) = a \,\}\Big)$$

  also called the *joint pmf*

Finite sample spaces

Suppose Ω is finite and $x : \Omega \to \mathbb{R}^2$. Then x_1 and x_2 only take a finite set of values.

$$x_1 : \Omega \to \{u_1, \dots, u_m\} \qquad\qquad x_2 : \Omega \to \{v_1, \dots, v_n\}$$

So we can represent the induced pmf by a matrix

$$\boxed{J_{ij} = p^x\left(\begin{bmatrix} u_i \\ v_j \end{bmatrix}\right)}$$

J is called the *joint probability matrix*

Marginal random variables

Suppose $x : \Omega \to \mathbb{R}^2$.

The *marginal random variables* are the components x_1 and x_2 of x.

The pmf of x_1 is

$$p^{x_1}(u_i) = \mathbf{Prob}(x_1 = u_i) = \mathbf{Prob}\big(\{\, \omega \in \Omega \mid x_1(\omega) = u_i \,\}\big) = \sum_{j=1}^{n} J_{ij}$$

So $p^{x_1} = J\mathbf{1}$ and similarly $p^{x_2} = J^T\mathbf{1}$

p^{x_1} and p^{x_2} are called the *marginal pmfs* of J

Independent random variables

If $x_1 : \Omega \to U$ and $x_2 : \Omega \to V$ are random variables, they are called *independent* if, for all $a \in U$ and $b \in V$,

$$\mathbf{Prob}(x_1 = a \text{ and } x_2 = b) = \mathbf{Prob}(x_1 = a)\,\mathbf{Prob}(x_2 = b)$$

Hence x_1 and x_2 are independent if and only if

$$J_{ij} = p_i^{x_1} p_j^{x_2}$$

because

$$\begin{aligned} J_{ij} &= \mathbf{Prob}(x_1 = u_i \text{ and } x_2 = v_j) \\ &= \mathbf{Prob}(x_1 = u_i)\,\mathbf{Prob}(x_2 = v_j) \\ &= p_i^{x_1} p_j^{x_2} \end{aligned}$$

the random variables x_1 and x_2 are independent if and only if $\mathbf{rank}(J) = 1$.

Independent random variables

We often have problems where

- $x_1 : \Omega \to \{1, \dots, m\}$ is a random variable
- $x_2 : \Omega \to \{1, \dots, n\}$ is a random variable
- x_1 and x_2 are independent

Here we *construct* the sample space

$$V = \Big\{ (i, j) \mid i \in \{1, \dots, m\} \text{ and } j \in \{1, \dots, n\} \Big\}$$

with pmf

$$p^x(i, j) = p^{x_1}(i)\, p^{x_2}(j)$$

The conditional pmf

Suppose $x : \Omega \to \mathbb{R}^2$ is a random vector with induced pmf $p^x : \mathbb{R}^2 \to [0,1]$.

Consider the conditional probability

$$\begin{aligned}\mathbf{Prob}(x_1 = a \mid x_2 = b) &= \mathbf{Prob}\Big(\{\,\omega \in \Omega \mid x_1(\omega) = a\,\} \,\Big|\, \{\,\omega \in \Omega \mid x_2(\omega) = b\,\}\Big) \\ &= \frac{\mathbf{Prob}\Big(\{\,\omega \in \Omega \mid x(\omega) = (a,b)\,\}\Big)}{\mathbf{Prob}\Big(\{\,\omega \in \Omega \mid x_2(\omega) = b\,\}\Big)} \\ &= \frac{p^x(a,b)}{p^{x_2}(b)}\end{aligned}$$

- Called the *conditional pmf of* x_1 *given* x_2
- Often written $p^{|x_2}(a,b)$

Example: four dice

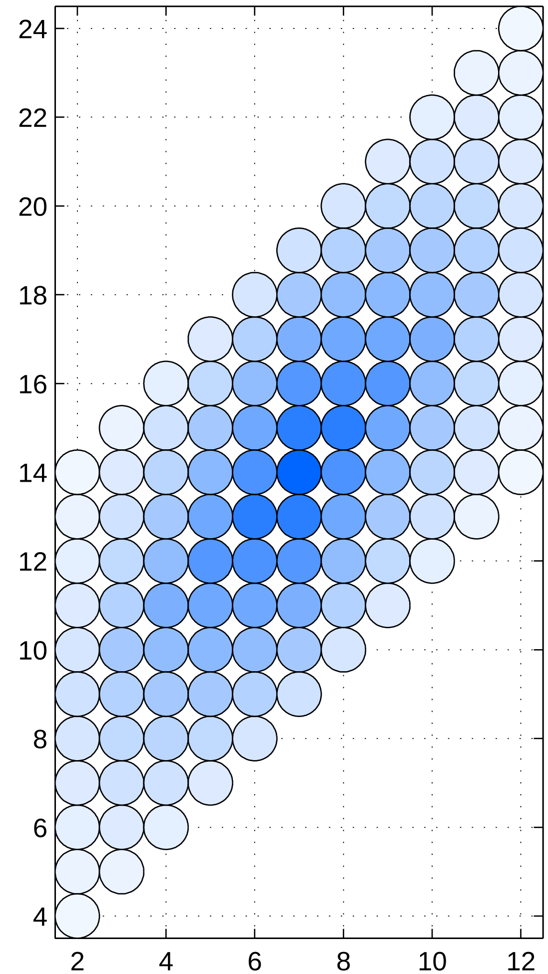
Four dice, sample space $\Omega = \Big\{ \omega \in \mathbb{R}^4 \mid \omega_i \in \{1, 2, \ldots, 6\} \Big\}$

- $x : \Omega \to \mathbb{R}$ is the sum of the first two dice

$$x = \omega_1 + \omega_2$$

- $y : \Omega \to \mathbb{R}$ is the sum of all four dice

$$y = \omega_1 + \omega_2 + \omega_3 + \omega_4$$



The induced pmf of $\begin{bmatrix} x \\ y \end{bmatrix}$ is

$$p^{x,y}(a, b) = \sum_{w \in \Omega \ \mid \ x(\omega)=a \text{ and } y(\omega)=b} p(\omega)$$

Example: marginal pmf

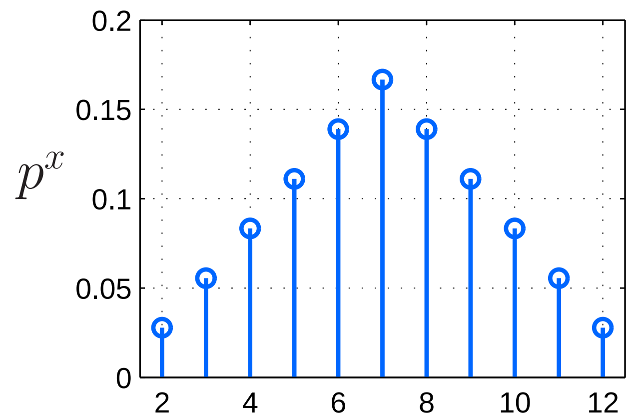
- The induced pmf of x is given by

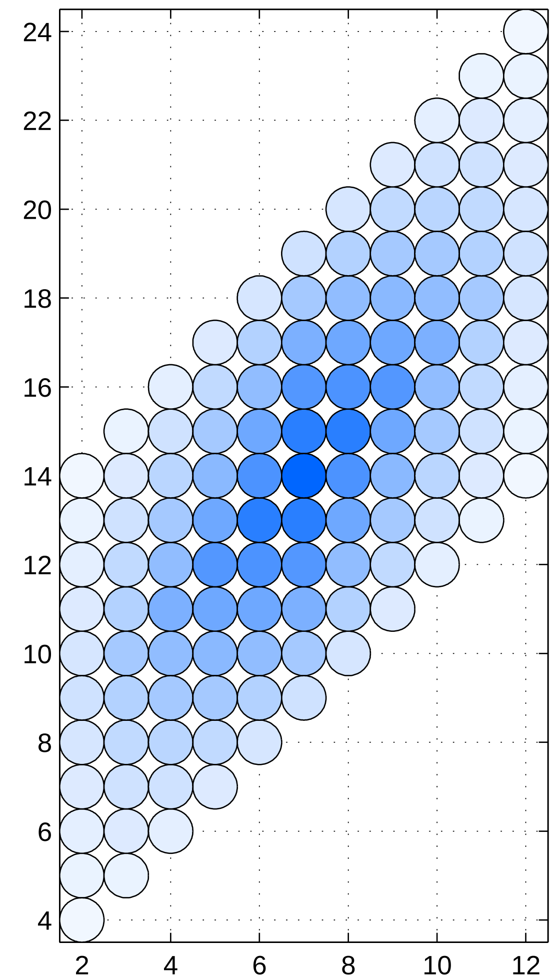
$$p^x(a) = \mathbf{Prob}(x = a)$$

- We can compute this from the induced pmf of $\begin{bmatrix} x \\ y \end{bmatrix}$ since

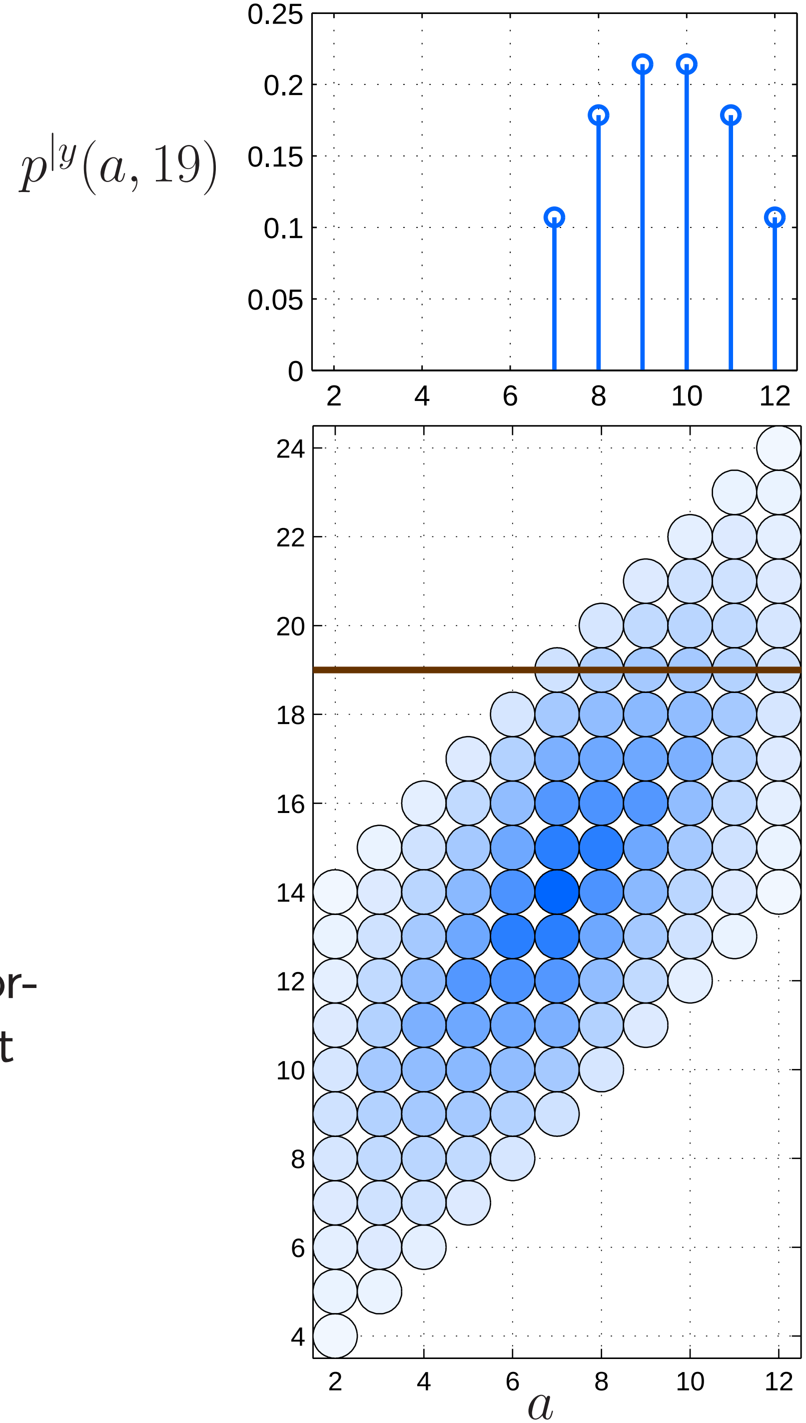
$$p^x(a) = \sum_{b=1}^{24} p^{x,y}(a, b)$$

- Also called the *prior* pmf of x, since it is the information we have about x before any measurements.





Example: conditional pmf



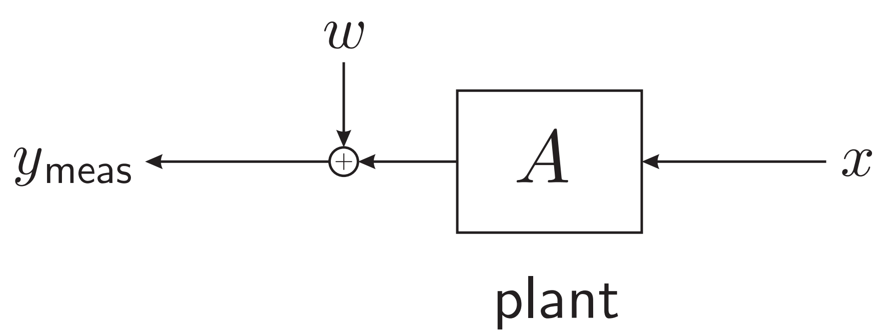
- Suppose we measure $y_{\text{meas}} = 19$

- The *conditional pmf* of x given $y = y_{\text{meas}}$ is shown.

- Also called the *posterior pmf* of x; i.e, the information we have regarding x after measurement

Example: dependent events

The above dice problem is a simple version of the following.



Here

- We would like to determine an estimate x_{est} of x, with 'good' properties
- We know a pmf for x (possibly obtained from previous measurements)
- w is random noise
- $y = Ax + w$ is *measured*

Since y and x are dependent, by measuring y we discover information about x

Covariance

Suppose $x : \Omega \to \mathbb{R}^n$. Let $\mu = \mathbf{E}\, x$ be the mean of x.

Define the *covariance* of x by

$$\boxed{\mathbf{cov}(x) = \mathbf{E}\big((x-\mu)(x-\mu)^T\big)}$$

- We'll often denote the covariance by $\Sigma = \mathbf{cov}(x)$
- Σ is symmetric and positive semidefinite because

$$\Sigma = \sum_{\omega \in \Omega} (x(\omega) - \mu)(x(\omega) - \mu)^T p(\omega)$$

- Σ_{ii} is the covariance of the i'th component x_i

$$\Sigma_{ii} = \mathbf{cov}(x_i)$$

Correlation

Let $\Sigma = \mathbf{cov}(x)$. The *correlation coefficient* of x_i and x_j is

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

- Since $\Sigma \geq 0$, we have $|\rho_{ij}| \leq 1$
- If $\rho_{ij} = 0$ then x_i and x_j are called *uncorrelated*.

Product of expectations

Suppose $x : \Omega \to \mathbb{R}^2$ is a 2-dimensional random variable and x_1 and x_2 are *independent*. Then

$$\mathbf{E}(x_1 x_2) = \mathbf{E}(x_1)\,\mathbf{E}(x_2)$$

To see this, suppose $x : \Omega \to \{u_1, \dots, u_m\} \times \{v_1, \dots, v_n\}$, then

$$\begin{aligned}
\mathbf{E}(x_1 x_2) &= \sum_{i=1}^{m} \sum_{j=1}^{n} p^x(u_i, v_j) u_i v_j \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} p_i^{x_1} p_j^{x_2} u_i v_j \\
&= \left(\sum_{i=1}^{m} p_i^{x_1} u_i \right) \left(\sum_{j=1}^{n} p_j^{x_2} v_j \right) \\
&= \mathbf{E}(x_1)\,\mathbf{E}(x_2)
\end{aligned}$$

Correlation and independence

Hence

> if x_i and x_j are independent, then they are uncorrelated

Because if x_i and x_j are independent, then $\Sigma_{ij} = 0$, since

$$\begin{aligned}\Sigma_{ij} &= \mathbf{E}(x_i - \mu_i)(x_j - \mu_j) \\ &= \mathbf{E}(x_i - \mu_i)\,\mathbf{E}(x_j - \mu_j) \\ &= 0\end{aligned}$$

- The converse is *not true*
- If $x : \Omega \to \mathbb{R}^n$, and for all i, j the random variables x_i and x_j are pairwise independent, then $\mathbf{cov}(x)$ is *diagonal*

Affine transformations of random vectors

Suppose $x : \Omega \to \mathbb{R}^n$ is a random variable, and $A \in \mathbb{R}^{m \times n}$. Let $y = Ax + b$. Then

- The mean of y is *the same affine function* of the mean of x

$$\boxed{\mathbf{E}\, y = A\, \mathbf{E}\, x + b}$$

- The covariance of y is a linear function of the covariance of x

$$\boxed{\mathbf{cov}(y) = A\, \mathbf{cov}(x) A^T}$$

These two facts will be very important in estimation for linear systems

Affine transformations of random vectors

For the mean, we have

$$\begin{aligned}
\mathbf{E}\,y &= \mathbf{E}(Ax+b) \\
&= \sum_{\omega\in\Omega}(Ax(\omega)+b)p(\omega) \\
&= b + A\sum_{\omega\in\Omega}x(\omega)p(\omega) \\
&= b + A\,\mathbf{E}\,x
\end{aligned}$$

And for the covariance

$$\begin{aligned}
\mathbf{cov}(y) &= \mathbf{E}\big((y-\mathbf{E}\,y)(y-\mathbf{E}\,y)^T\big) \\
&= \mathbf{E}\big(A(x-\mathbf{E}\,x)(x-\mathbf{E}\,x)^T A^T\big) \\
&= A\,\mathbf{E}\big((x-\mathbf{E}\,x)(x-\mathbf{E}\,x)^T\big)A^T \\
&= A\,\mathbf{cov}(x)A^T
\end{aligned}$$

Sum of two independent random variables

Suppose x and y are *independent* random variables which take *integer values*

- $x : \Omega \to \{0, 1, \dots, m\}$
- $y : \Omega \to \{0, 1, \dots, n\}$

Define the random variable $z : \Omega \to \{0, \dots, m+n\}$ by

$$z = x + y$$

The induced pmf of z is the *convolution* of p^x and p^y.

$$p^z_k = \sum_{i=0}^{k} p^x_i p^y_{k-i}$$

Sum of two independent random variables

The proof is as follows.

$$\begin{aligned} p_k^z &= \mathbf{Prob}(z = k) \\ &= \sum \Big\{ \mathbf{Prob}(x = i \text{ and } y = j) \mid i + j = k \Big\} \\ &= \sum \Big\{ p_i^x p_j^y \mid i + j = k \Big\} \\ &= \sum_{i=0}^{k} p_i^x p_{k-i}^y \end{aligned}$$

IID random variables

Suppose $x_1, \dots, x_n$ are random variables, with $x_i : \Omega \to \mathbb{R}^m$

- They are called *identically distributed* if each x_i has the same pmf.
- If they are also *independent*, they are called *IID*.

Suppose $x_1, \dots, x_n$ are IID, each with mean v and covariance Q. Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then the mean and covariance of x is

$$\mathbf{E}\, x = \begin{bmatrix} v \\ v \\ \vdots \\ v \end{bmatrix} \qquad \mathbf{cov}(x) = \begin{bmatrix} Q & & & \\ & Q & & \\ & & \ddots & \\ & & & Q \end{bmatrix}$$

IID random variables

Define the average $s_n : \Omega \to \mathbb{R}^m$

$$s_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

We have $s_n = Ax$ where $A = \frac{1}{n} \begin{bmatrix} I & I & \dots & I \end{bmatrix}$, so its mean is

$$\mathbf{E}\, s_n = A\, \mathbf{E}\, x = A \begin{bmatrix} v \\ v \\ \vdots \\ v \end{bmatrix} = v$$

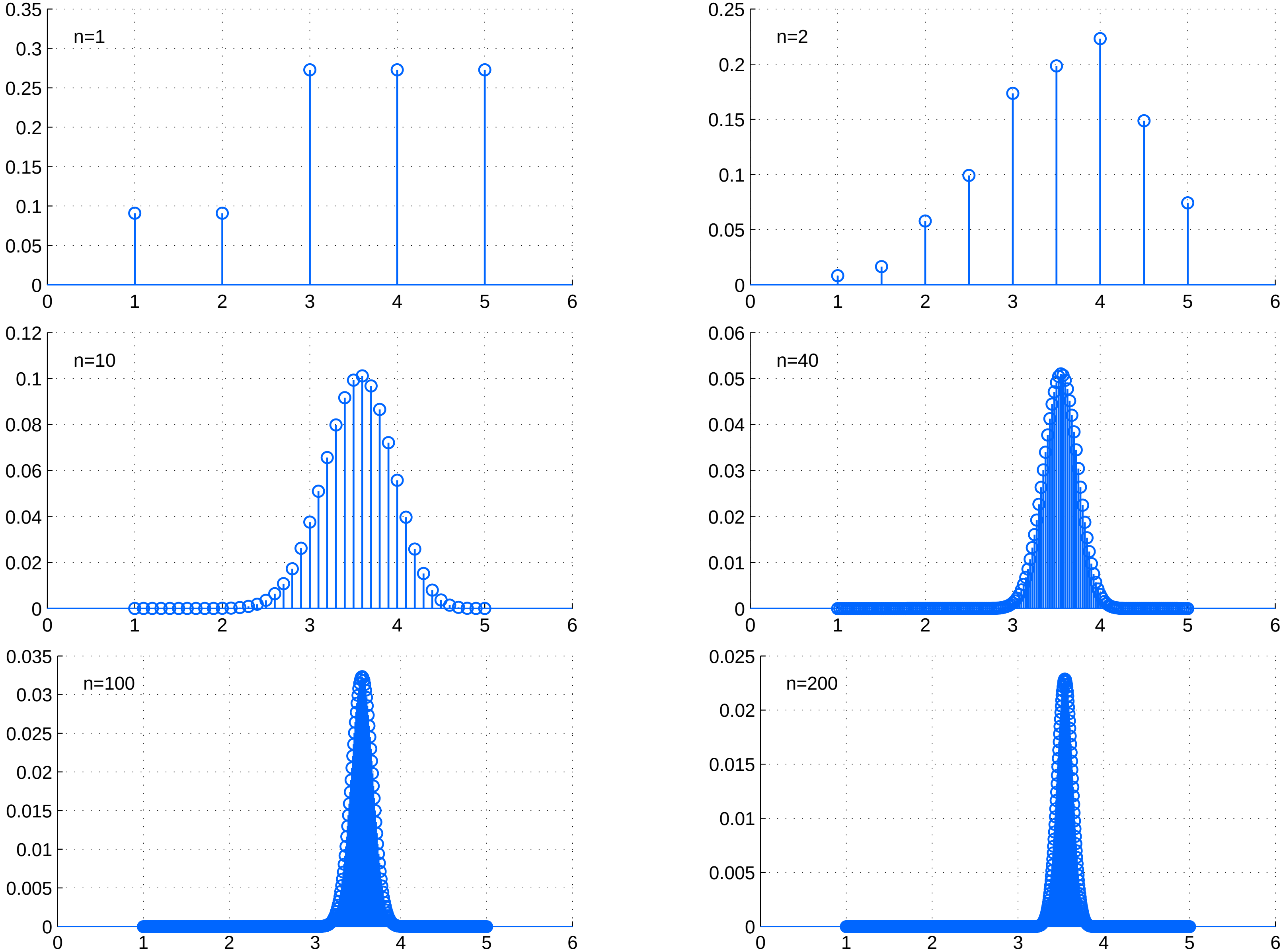
and its covariance is

$$\mathbf{cov}(s_n) = A\, \mathbf{cov}(x) A^T = A \begin{bmatrix} Q & & & \\ & Q & & \\ & & \ddots & \\ & & & Q \end{bmatrix} A^T = \frac{Q}{n}$$

so taking the average of n IID random variables *reduces the covariance* by a factor of n

Example: sums of IID random variables

The pmf of s_n for various values of n



Law of large numbers

Suppose $x_i : \Omega \to \mathbb{R}$ are *real-valued* and IID, each with mean v and variance Q. The (weak) *law of large numbers* is

$$\lim_{n\to\infty} \mathbf{Prob}\big(|s_n - v| \leq \varepsilon\big) = 1$$

Proof is just from the the Chebyshev inequality

$$\mathbf{Prob}\big(|s_n - v| \leq \varepsilon\big) \geq 1 - \frac{Q}{n\varepsilon^2}$$

- Here s_n is the average of n IID random variables.
- The law of large numbers says that the probability that s_n is within ε of the mean tends to 1 as n becomes large.
- We say the sequence of random variables $s_0, s_1, \ldots$ *converges in probability* to v.

Law of large numbers

We have the Chebyshev inequality

$$\mathbf{Prob}\big(|s_n - v| \leq \varepsilon\big) \geq 1 - \frac{Q}{n\varepsilon^2}$$

Hence to achieve confidence width of ε at probability p_{conf}, we need

$$\varepsilon = \sqrt{\frac{Q}{n(1 - p_{\mathsf{conf}})}}$$

i.e., *the confidence width* decreases as $\dfrac{1}{\sqrt{n}}$.

Estimating the mean

If we perform *identical repeated experiments*, then for large n the *sample mean*

$$s_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

will be close to the true mean with high probability.

Hence we can use this to *estimate* the mean of random data.

Identifying a pmf from data

We have random samples $x_1, x_2, \ldots$ of $x : \Omega \to V$, and we would like to estimate p^x

- Define the indicator random variable $I_j : V \to \mathbb{R}$ by

$$I_a(y) = \begin{cases} 1 & \text{if } y = a \\ 0 & \text{otherwise} \end{cases}$$

  The expected value of I_a is

$$\mathbf{E}\, I_a = \sum_{b \in V} p^x(b) I_a(b) = p^x(a)$$

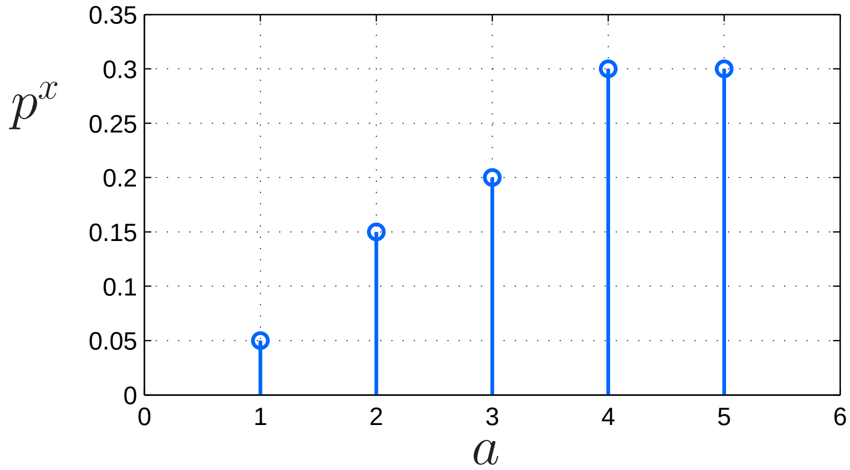
- So to estimate $p^x(a)$, we compute the sample mean

$$s(a, n) = \frac{1}{n} \sum_{i=1}^{n} I_a(x_i)$$

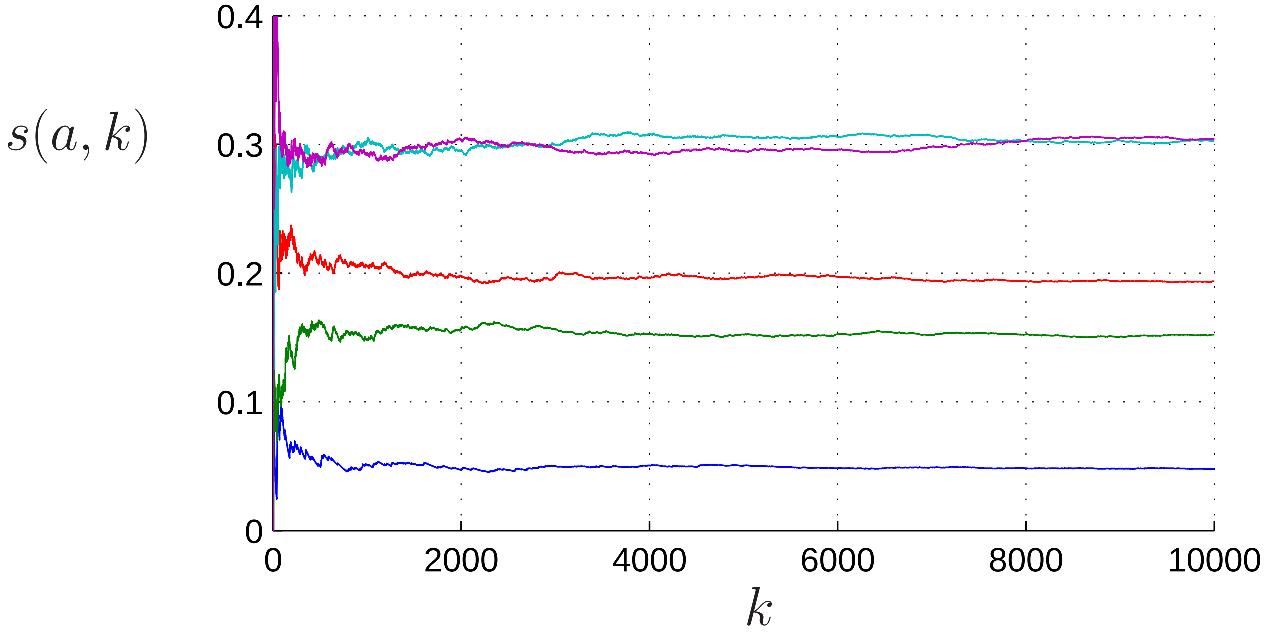
  which is exactly the *relative frequency* of outcome $x = a$ in n trials.

Example: identifying a pmf

We have the pmf

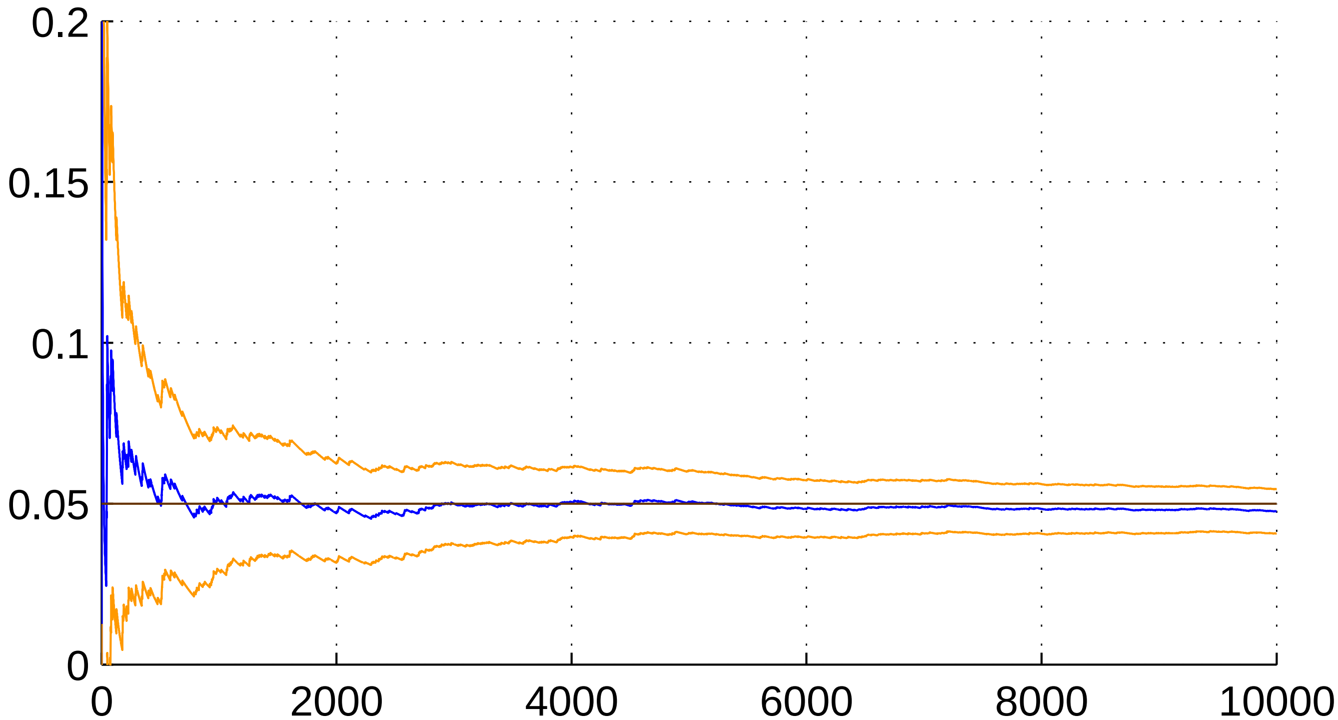


Simulating gives relative frequencies $s(a,k)$ plotted against k for different a



Example: identifying a pmf

The graph shows the relative frequency $s(n, 1)$ and its confidence intervals



The orange curves give the 90% confidence interval

$$s(a, k) \pm \sqrt{\frac{\mathbf{cov}(I_a)}{k(1 - p_{\text{conf}})}}$$