Continuing studies in Probability

Last lecture of Probability IA, 2015

Richard Weber

Topics to be revisited

- Markov, Chebyshev and Jensen's inequalities,
- moment generating function,
- sums of Bernoulli r.vs,
- Stirling's formula,
- normal distribution,
- gambler's ruin,
- tower property of conditional expectation,
- expected value of a sum of r.vs,
- Dyke words,
- generating functions,
- change of variable,
- convergence in distribution,
- Central limit theorem.

Games of chance: American roulette wheel

The wheel has 38 slots: 18 red, 18 black, 0 and 00.

Betting on red has probability of success
$p = 18/38 = 0.4737$.

Large deviations and Chernoff's bound

Gambler's success

John plays roulette at Las Vegas, placing $\$1$ on red at each turn, which is then doubled, with probability $p = \frac{9}{19}$, or lost, with probability $q = 1 - p = \frac{10}{19}$.

He tries to increase his wealth by $\$100$.

Gambler's success

John plays roulette at Las Vegas, placing $\$1$ on red at each turn, which is then doubled, with probability $p = \frac{9}{19}$, or lost, with probability $q = 1 - p = \frac{10}{19}$.

He tries to increase his wealth by $\$100$.

This is very unlikely. The probability he should ever be up by $\$100$ is $(p/q)^{100} = (9/10)^{100} = 0.0000265$.

Gambler's success

John plays roulette at Las Vegas, placing $\$1$ on red at each turn, which is then doubled, with probability $p = \frac{9}{19}$, or lost, with probability $q = 1 - p = \frac{10}{19}$.

He tries to increase his wealth by $\$100$.

This is very unlikely. The probability he should ever be up by $\$100$ is $(p/q)^{100} = (9/10)^{100} = 0.0000265$.

But suppose this rare event or **large deviation** occurs and after n games he is up by $\$100$.

What can we say about n, and about the path followed by John's wealth?

Model for the game

$X_1, \ldots, X_n$ are i.i.d. $B(1, p)$.
$S_n = X_1 + \cdots + X_n$ be the number of games John wins in the first n.

His wealth is $W_n = 2S_n - n$.

Let $\mu = EX_1 = p$ and $\sigma^2 = \mathrm{Var}(X_i) = pq$.

Model for the game

$X_1, \ldots, X_n$ are i.i.d. $B(1, p)$.
$S_n = X_1 + \cdots + X_n$ be the number of games John wins in the first n.

His wealth is $W_n = 2S_n - n$.

Let $\mu = EX_1 = p$ and $\sigma^2 = \mathrm{Var}(X_i) = pq$.

Note that $EW_n = -(q-p)n < 0$.

To reach 100 he must win $\frac{1}{2}n + 50$, and lose $\frac{1}{2}n - 50$, games.

Some bounds

How likely is $P(S_n > na)$ for $a > p$?

Some bounds

How likely is $P(S_n > na)$ for $a > p$?

Using the Chebyshev bound

$$P(S_n > na) = P\left(S_n - n\mu \geq n(a-\mu)\right) \leq \frac{\mathsf{Var}(S_n)}{n^2(a-\mu)^2} = \frac{\sigma^2}{n(a-\mu)^2}.$$

Some bounds

How likely is $P(S_n > na)$ for $a > p$?

Using the Chebyshev bound

$$P(S_n > na) = P\left(S_n - n\mu \geq n(a-\mu)\right) \leq \frac{\mathsf{Var}(S_n)}{n^2(a-\mu)^2} = \frac{\sigma^2}{n(a-\mu)^2}.$$

Alternatively, by the Central limit theorem,

$$P(S_n > na)) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} > \frac{(a-\mu)\sqrt{n}}{\sigma}\right) \approx 1 - \Phi\left(\frac{(a-\mu)\sqrt{n}}{\sigma}\right).$$

Both show that $P(S_n > na) \to 0$ as $n \to \infty$.

Chernoff upper bound

Let $m(\theta) = Ee^{\theta X_1}$ be the m.g.f. of X_1. Let $\theta > 0$,

$$P(S_n > na) = P\left(e^{\theta S_n} > e^{\theta na}\right)$$

Chernoff upper bound

Let $m(\theta) = Ee^{\theta X_1}$ be the m.g.f. of X_1. Let $\theta > 0$,

$$P(S_n > na) = P\Big(e^{\theta S_n} > e^{\theta na}\Big)$$

$$\leq \frac{E[e^{\theta S_n}]}{e^{\theta na}} \quad \text{(by Markov inequality)}$$

Chernoff upper bound

Let $m(\theta) = Ee^{\theta X_1}$ be the m.g.f. of X_1. Let $\theta > 0$,

$$P(S_n > na) = P\left(e^{\theta S_n} > e^{\theta na}\right)$$

$$\leq \frac{E[e^{\theta S_n}]}{e^{\theta na}} \quad \text{(by Markov inequality)}$$

$$= \left(\frac{m(\theta)}{e^{\theta a}}\right)^n$$

Chernoff upper bound

Let $m(\theta) = Ee^{\theta X_1}$ be the m.g.f. of X_1. Let $\theta > 0$,

$$
\begin{aligned}
P(S_n > na) &= P\left(e^{\theta S_n} > e^{\theta na}\right) \\
&\leq \frac{E[e^{\theta S_n}]}{e^{\theta na}} \quad \text{(by Markov inequality)} \\
&= \left(\frac{m(\theta)}{e^{\theta a}}\right)^n \\
&= e^{-n[\theta a - \log m(\theta)]}.
\end{aligned}
$$

Chernoff upper bound

Let $m(\theta) = Ee^{\theta X_1}$ be the m.g.f. of X_1. Let $\theta > 0$,

$$P(S_n > na) \leq e^{-n[\theta a - \log m(\theta)]}.$$

Chernoff upper bound

Let $m(\theta) = Ee^{\theta X_1}$ be the m.g.f. of X_1. Let $\theta > 0$,

$$P(S_n > na) \le e^{-n[\theta a - \log m(\theta)]}.$$

Now minimize rhs over θ to get the best bound.

$$P(S_n > na) \le e^{-nI(a)} \quad \text{(the \textbf{Chernoff bound})}, \quad (1)$$

where $I(a) = \max_{\theta>0} [\theta a - \log m(\theta)]$.

Chernoff lower bound

$$P(S_n > na) \leq e^{-nI(a)} \quad \text{(the \textbf{Chernoff bound})}, \quad (1)$$

where $I(a) = \max_{\theta>0}[\theta a - \log m(\theta)]$.

Chernoff lower bound

$$P(S_n > na) \le e^{-nI(a)} \quad \text{(the \textbf{Chernoff bound})}, \quad (1)$$

where $I(a) = \max_{\theta>0}[\theta a - \log m(\theta)]$.

This bound is tight, in that given any $\delta > 0$,

$$P(S_n > na) \ge e^{-n(I(a)+\delta)}, \quad (2)$$

for all sufficiently large n.

Chernoff lower bound

$$P(S_n > na) \le e^{-nI(a)} \quad \text{(the \textbf{Chernoff bound})}, \quad (1)$$

where $I(a) = \max_{\theta>0}[\theta a - \log m(\theta)]$.

This bound is tight, in that given any $\delta > 0$,

$$P(S_n > na) \ge e^{-n(I(a)+\delta)}, \qquad (2)$$

for all sufficiently large n. Hence

$$\log P(S_n > an) \sim -nI(a) \qquad (3)$$

$\sim$ means that the quotient of the two sides tends to 1 as $n \to \infty$.

Case of normal random variable

This holds for random variables more generally. If $X_i \sim N(0,1)$ then $m(\theta) = e^{\frac{1}{2}\theta^2}$ and

$$I(a) = \max_\theta \big[\theta a - \tfrac{1}{2}\theta^2\big] = \tfrac{1}{2}a^2.$$

So

$$\log P(S_n > an) \sim -n\tfrac{1}{2}a^2.$$

Case of normal random variable

$$\log P(S_n > an) \sim -n\frac{1}{2}a^2.$$

Case of normal random variable

$$\log P(S_n > an) \sim -n\tfrac{1}{2}a^2.$$

Compare this to

$$\begin{aligned}
P(S_n > an) &= \int_{an}^{\infty} \frac{1}{\sqrt{2\pi n}} e^{-t^2/(2n)} dt \\
&< \int_{an}^{\infty} \frac{1}{\sqrt{2\pi n}} \left(1 + \frac{n}{t^2}\right) e^{-t^2/(2n)} dt \\
&= \frac{1}{\sqrt{2\pi a^2 n}} e^{-n\frac{1}{2}a^2}.
\end{aligned}$$

Herman and Judy Chernoff

Case of Bernoulli random variable

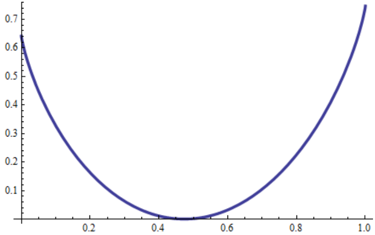
For $B(1, p)$ the m.g.f. is $m(\theta) = q + pe^{\theta}$

Case of Bernoulli random variable

For $B(1,p)$ the m.g.f. is $m(\theta) = q + pe^{\theta}$ and

$$I(a) = \max_{\theta}[\theta a - m(\theta)] = (1-a)\log\frac{1-a}{1-p} + a\log\frac{a}{p},$$

which is convex in a, with its minimum $I(p) = 0$.



$I(a)$ for $p = 18/38$.

Case of Bernoulli random variable

For $B(1,p)$ the m.g.f. is $m(\theta) = q + pe^{\theta}$ and

$$I(a) = \max_{\theta}[\theta a - m(\theta)] = (1-a)\log\frac{1-a}{1-p} + a\log\frac{a}{p},$$

which is convex in a, with its minimum $I(p) = 0$.

Can also verify the lower bound (2). Let $j_n = \lceil na \rceil$.

$$P(S_n > na) = \sum_{i \geq j_n}^{n} \binom{n}{i} p^i (1-p)^{n-i} > \binom{n}{j_n} p^{j_n} (1-p)^{n-j_n}.$$

Applying Stirling's formula on the rhs we will find:

$$\underline{\lim}_{n\to\infty}(1/n)\log P(S_n > na) = -I(a).$$

Hence $\log P(S_n > an) \sim -nI(a)$.

Most likely path to $\$100$.

The path on which John's wealth increases to $\$100$ is most likely to look like a straight line.

Most likely path to $\$100$.

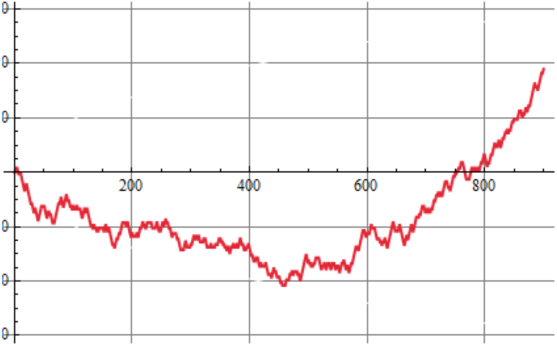
The path on which John's wealth increases to $\$100$ is most likely to look like a straight line.

For instance, suppose S_n increases at rate a_1 for n_1 bets, and then rate a_2 for n_2 bets, where $n_1 + n_2 = n$ and $2(n_1 a_1 + n_2 a_2) - n = 100$.

Most likely path to $\$100$.

The path on which John's wealth increases to $\$100$ is most likely to look like a straight line.

For instance, suppose S_n increases at rate a_1 for n_1 bets, and then rate a_2 for n_2 bets, where $n_1 + n_2 = n$ and $2(n_1 a_1 + n_2 a_2) - n = 100$.



Most likely path to $\$100$.

The path on which John's wealth increases to $\$100$ is most likely to look like a straight line.

For instance, suppose S_n increases at rate a_1 for n_1 bets, and then rate a_2 for n_2 bets, where $n_1 + n_2 = n$ and $2(n_1 a_1 + n_2 a_2) - n = 100$.

The log-probability of this is about $-n_1 I(a_1) - n_2 I(a_2)$, which is maximized by $a_1 = a_2$, since I is a convex function.

$$-\frac{n_1 I(a_1) + n_2 I(a_2)}{n_1 + n_2} \leq -I\left(\frac{n_1 a_1 + n_2 a_2}{n_1 + n_2}\right)$$

by Jensen's inequality, with equality when $a_1 = a_2$.

Most likely n to $\$100$.

So the most likely route to 100 is over n bets, with S_n increasing at a constant rate a, and such that $2na - n = 100$.

Subject to these constraints $\log P(S_n > an) \approx -nI(a)$ is maximized by $n = 100/(1-2p)$, $a = 1 - 2p$.

Most likely n to $\$100$.

So the most likely route to 100 is over n bets, with S_n increasing at a constant rate a, and such that $2na - n = 100$.

Subject to these constraints
$\log P(S_n > an) \approx -nI(a)$ is maximized by
$n = 100/(1-2p)$, $a = 1-2p$.
This means it is highly likely that

$$n \approx 100/(1 - 2 \times (18/38)) = 1900.$$

Interestingly, this is the same as the number of games over which his expected loss would be $\$100$.

Other applications

Recall from Examples sheet 1,

16. Suppose a die is rolled n times. Show that the probability of a roll of i appearing k_i times, $i = 1, \ldots, 6$, is

$$\phi(k_1, \ldots, k_6) = \frac{n!}{k_1! \cdots k_6!} \frac{1}{6^n}.$$

The expected value of the total of n rolls is $3.5n$. Suppose $\sum_i ik_i = \rho n$, $1 \leq \rho \leq 6$, and $k_i = np_i$. Use Stirling's formula to show that subject to these constraints, ϕ is maximized by choosing the p_i as nonnegative numbers that solve the optimization problem

$$\underset{p_1,\ldots,p_6}{\text{maximize}} - \sum_i p_i \log p_i, \quad \text{subject to } \sum_i p_i = 1, \text{ and } \sum_i ip_i = \rho.$$

For $\rho = 4$ the maximizer is $p^* \approx (0.103, 0.123, 0.146, 0.174, 0.207, 0.247)$, so the most likely way of obtaining a total of $4n$ is when about 24.7% of the dice rolls are a 6. What do you guess would be the solution to the optimization problem if $\rho = 3.5$? If $\rho = 3$?

There are many other applications of large deviations theory, especially in coding, and queueing.

Random matrices

Random matrices

Consider a symmetric $n \times n$ matrix A, constructed by setting diagonal elements 0, and independently choosing each off-diagonal $a_{ij} = a_{ji}$ as 1 or -1 by tossing a fair coin.

A random 10×10 symmetric matrix, having eigenvalues -4.515, -4.264, -2.667, -1.345, -0.7234, 1.169, 2.162, 2.626, 3.279, 4.277.

$$\begin{pmatrix} 0 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & 0 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 0 & -1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 0 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 0 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 0 & 1 & -1 \\ -1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & 0 & -1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 0 \end{pmatrix}$$

Recall that the eigenvalues of a symmetric real matrix are real.

Moments of a random eigenvalue

Let Λ be a randomly chosen eigenvalue of a random A. *What can we say about the r.v.* Λ*?*

Moments of a random eigenvalue

Let Λ be a randomly chosen eigenvalue of a random A. *What can we say about the r.v.* Λ*?*

A and $-A$ are equally likely, so $E\Lambda^k = 0$ if k is odd.

Moments of a random eigenvalue

Let Λ be a randomly chosen eigenvalue of a random A. *What can we say about the r.v. Λ?*

A and $-A$ are equally likely, so $E\Lambda^k = 0$ if k is odd.

Consider $k = 4$. Eigenvalues of A are $\lambda_1, \ldots, \lambda_n$.

$$E[\Lambda^4] = \frac{1}{n}E[\lambda_1^4 + \cdots + \lambda_n^4] = \frac{1}{n}E[\mathrm{Tr}(A^4)]$$

(4)

Moments of a random eigenvalue

Let Λ be a randomly chosen eigenvalue of a random A. *What can we say about the r.v. Λ?*

A and $-A$ are equally likely, so $E\Lambda^k = 0$ if k is odd.

Consider $k = 4$. Eigenvalues of A are $\lambda_1, \ldots, \lambda_n$.

$$\begin{aligned} E[\Lambda^4] &= \tfrac{1}{n}E[\lambda_1^4 + \cdots + \lambda_n^4] = \tfrac{1}{n}E[\mathsf{Tr}(A^4)] \\ &= \tfrac{1}{n}E\sum_{i_1,i_2,i_3,i_4} a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1} \end{aligned} \tag{4}$$

Moments of a random eigenvalue

Let Λ be a randomly chosen eigenvalue of a random A. *What can we say about the r.v. Λ?*

A and $-A$ are equally likely, so $E\Lambda^k = 0$ if k is odd.

Consider $k = 4$. Eigenvalues of A are $\lambda_1, \ldots, \lambda_n$.

$$\begin{aligned} E[\Lambda^4] &= \tfrac{1}{n}E[\lambda_1^4 + \cdots + \lambda_n^4] = \tfrac{1}{n}E[\mathrm{Tr}(A^4)] \\ &= \tfrac{1}{n}E \sum_{i_1,i_2,i_3,i_4} a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1} \\ &= \tfrac{1}{n} \sum_{i_1,i_2,i_3,i_4} E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] \end{aligned} \tag{4}$$

taking the sum over all paths of length 4 through a subset of the n indices: $i_1 \to i_2 \to i_3 \to i_4 \to i_1$.

Limits for moments of a random eigenvalue

$$E[\mathrm{Tr}(A^4)] = \sum_{i_1,i_2,i_3,i_4} E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] \quad (4)$$

$E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] = 1$ if for each two i, j the total number of a_{ij}s and a_{ji}s in $\{a_{i_1 i_2}, a_{i_2 i_3}, a_{i_3 i_4}, a_{i_4 i_1}\}$ is even. This happens for

- $n(n-1)(n-2)$ terms of the form $E[a_{ij} a_{ji} a_{ik} a_{ki}]$;

Limits for moments of a random eigenvalue

$$E[\mathrm{Tr}(A^4)] = \sum_{i_1,i_2,i_3,i_4} E[a_{i_1i_2}a_{i_2i_3}a_{i_3i_4}a_{i_4i_1}] \qquad (4)$$

$E[a_{i_1i_2}a_{i_2i_3}a_{i_3i_4}a_{i_4i_1}] = 1$ if for each two i, j the total number of a_{ij}s and a_{ji}s in $\{a_{i_1i_2}, a_{i_2i_3}, a_{i_3i_4}, a_{i_4i_1}\}$ is even. This happens for

- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{ji}a_{ik}a_{ki}]$;
- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{jk}a_{kj}a_{ji}]$;

Limits for moments of a random eigenvalue

$$E[\mathrm{Tr}(A^4)] = \sum_{i_1,i_2,i_3,i_4} E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] \qquad (4)$$

$E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] = 1$ if for each two i, j the total number of a_{ij}s and a_{ji}s in $\{a_{i_1 i_2}, a_{i_2 i_3}, a_{i_3 i_4}, a_{i_4 i_1}\}$ is even. This happens for

- $n(n-1)(n-2)$ terms of the form $E[a_{ij} a_{ji} a_{ik} a_{ki}]$;
- $n(n-1)(n-2)$ terms of the form $E[a_{ij} a_{jk} a_{kj} a_{ji}]$;
- $n(n-1)$ terms of the form $E[a_{ij} a_{ji} a_{ij} a_{ji}]$.

Limits for moments of a random eigenvalue

$$E[\mathrm{Tr}(A^4)] = \sum_{i_1,i_2,i_3,i_4} E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] \qquad (4)$$

$E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] = 1$ if for each two i, j the total number of a_{ij}s and a_{ji}s in $\{a_{i_1 i_2}, a_{i_2 i_3}, a_{i_3 i_4}, a_{i_4 i_1}\}$ is even. This happens for

- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{ji}a_{ik}a_{ki}]$;
- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{jk}a_{kj}a_{ji}]$;
- $n(n-1)$ terms of the form $E[a_{ij}a_{ji}a_{ij}a_{ji}]$.

Otherwise $E[a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} a_{i_4 i_1}] = 0$.

Limits for moments of a random eigenvalue

$$E[\mathsf{Tr}(A^4)] = \sum_{i_1,i_2,i_3,i_4} E[a_{i_1i_2}a_{i_2i_3}a_{i_3i_4}a_{i_4i_1}] \qquad (4)$$

$E[a_{i_1i_2}a_{i_2i_3}a_{i_3i_4}a_{i_4i_1}] = 1$ if for each two i, j the total number of a_{ij}s and a_{ji}s in $\{a_{i_1i_2}, a_{i_2i_3}, a_{i_3i_4}, a_{i_4i_1}\}$ is even. This happens for

- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{ji}a_{ik}a_{ki}]$;
- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{jk}a_{kj}a_{ji}]$;
- $n(n-1)$ terms of the form $E[a_{ij}a_{ji}a_{ij}a_{ji}]$.

Otherwise $E[a_{i_1i_2}a_{i_2i_3}a_{i_3i_4}a_{i_4i_1}] = 0$.Thus,

$$E[\Lambda^4/n^{\frac{4}{2}}] = n^{-\frac{4}{2}-1}E[\mathsf{Tr}(A^4)]$$
$$= n^{-3}[2n(n-1)(n-2) + n(n-1)] \to 2, \text{ as } n \to \infty.$$

Dyck words, again!

- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{ji}a_{ik}a_{ki}]$;
- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{jk}a_{kj}a_{ji}]$;
- $n(n-1)$ terms of the form $E[a_{ij}a_{ji}a_{ij}a_{ji}]$.

$$E[\Lambda^4/n^{\frac{4}{2}}] = n^{-3}E[\mathrm{Tr}(A^4)] \to 2, \text{ as } n \to \infty.$$

The limit 2 is C_2, the number of Dyke words of length 4.

These words are ()() and (()), matching the patterns of the first two bullet points above.

Dyck words, again!

- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{ji}a_{ik}a_{ki}]$;
- $n(n-1)(n-2)$ terms of the form $E[a_{ij}a_{jk}a_{kj}a_{ji}]$;
- $n(n-1)$ terms of the form $E[a_{ij}a_{ji}a_{ij}a_{ji}]$.

$$E[\Lambda^4/n^{\frac{4}{2}}] = n^{-3}E[\mathsf{Tr}(A^4)] \to 2, \text{ as } n \to \infty.$$

The limit 2 is C_2, the number of Dyke words of length 4.

These words are ()() and (()), matching the patterns of the first two bullet points above.

Generalizes to any even k, to show
$\lim_{n\to\infty} E[(\Lambda/n^{\frac{1}{2}})^k] \to C_{k/2}$,

a Catalan number, and the number of Dyck words of length k (described in §12.2).

Semicircle p.d.f.

This begs the question: what random variable X has sequence of moments

$$\{EX^k\}_{k=1}^{\infty} = \{0, C_1, 0, C_2, 0, C_3, \dots\} = \{0, 1, 0, 2, 0, 5, \dots\}?$$

Semicircle p.d.f.

This begs the question: what random variable X has sequence of moments

$\{EX^k\}_{k=1}^{\infty} = \{0, C_1, 0, C_2, 0, C_3, \dots\} = \{0, 1, 0, 2, 0, 5, \dots\}$?

It is easy to check that this is true when X has the p.d.f.

$$f(x) = \tfrac{1}{2\pi}\sqrt{4 - x^2}, \quad -2 \le x \le 2.$$

Semicircle p.d.f.

This begs the question: what random variable X has sequence of moments

$\{EX^k\}_{k=1}^{\infty} = \{0, C_1, 0, C_2, 0, C_3, \dots\} = \{0, 1, 0, 2, 0, 5, \dots\}$?

It is easy to check that this is true when X has the p.d.f.

$$f(x) = \tfrac{1}{2\pi}\sqrt{4 - x^2}, \quad -2 \le x \le 2.$$

$$C_k = \int_{-2}^{2} (x^2)^k \tfrac{1}{2\pi}\sqrt{4 - x^2}\, dx,$$

which is related to the fact that the generating function for the Catalan numbers is

$$c(x) = \sum_{k=0}^{\infty} C_k x^k = \frac{1 - \sqrt{1 - 4x}}{2x} = \sum_{n=0}^{\infty} \binom{2k}{k} \frac{x^k}{k+1}.$$

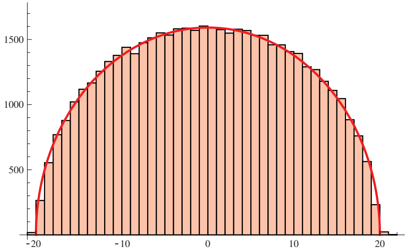
Wigner's semicircle law

Histogram of 50,000 eigenvalues obtained by randomly generating 500 random 100×100 matrices.
Bin sizes are of width 1.

$\lambda/\sqrt{100}$ has empirical density closely matching f.

Rescaling appropriately,
the red semicircle is

$$g(x) = 50000\tfrac{1}{10}f(\tfrac{x}{10}), \quad -20 \le x \le 20.$$



This result is **Wigner's semicircle law**:

$$\Lambda/\sqrt{n} \to_{\mathscr{D}} \text{ r.v. with semicircle p.d.f. } f.$$

Wigner's semicircle law is a universal law

Notice we did not really need the assumption that a_{ij} are chosen from the discrete uniform distribution on $\{-1, 1\}$.

Need only that $Ea_{ij}^k = 0$ for odd k and $Ea_{ij}^k < \infty$ for even k.

Wigner's semicircle law is a universal law

Notice we did not really need the assumption that a_{ij} are chosen from the discrete uniform distribution on $\{-1, 1\}$.

Need only that $Ea_{ij}^k = 0$ for odd k and $Ea_{ij}^k < \infty$ for even k.

Wigner's theorem is in the same spirit as the Central limit theorem, which holds for any random variable with finite first two moments.

Wigner's theorem dates from 1955, but the finer analysis of the eigenvalues structure of random matrices interests researchers in the present day.

Topics revisted

Large deviations and random matrices are fruits of research in probability in modern times.

In looking at them we have touched on many topics covered in our course:

- Markov, Chebyshev and Jensen's inequalities,
- moment generating function,
- sums of Bernoulli r.vs,
- Stirling's formula,
- normal distribution,
- gambler's ruin,
- tower property of conditional expectation,
- expected value of a sum of r.vs,
- Dyke words,
- generating functions,
- change of variable,
- convergence in distribution,
- Central limit theorem.

Applicable courses in IB

Statistics

Statistics addresses the question,

"**What is this data telling me?**"

How should we design experiments and interpret their results?

In the Probability IA course we had the weak law of large numbers. This underlies the frequentist approach of estimating the probability p with which a drawing pin lands "point up" by tossing it many times and looking at the proportion of landings "point up". Bayes Theorem also enters into statistics.

To address questions about **estimation and hypothesis testing** we must model uncertainty and the way data is arises. That gives Probability a central role in Statistics. In the Statistics IB course you will put to good use what you have learned about random variables and distributions this year.

Markov chains

A Markov Chain is a generalization of the idea of a sequence of i.i.d. r.vs., $X_1, X_2, \dots$. There is a departure from independence because we now allow the distribution of X_{n+1} to depend on the value of X_n.

Many things in the world are like this: e.g. tomorrow's weather state follows in some random way from today's weather state.

A **random walk is a Markov chain**, as we have met in Probability IA.

In Markov Chains IB you will learn many more interesting things about random walks.

For example, Polya's theorem about random walk implies that it is possible to play the clarinet in our 3-D world but that this would be impossible in 2-D Flatland.

Optimization

Randomizing strategies are used in two-person games.

Consider game scissors-stone-paper, for which the optimal strategy is to randomize with probabilities $1/3$, $1/3$, $1/3$.

In the Optimization course you will learn how to solve other games.

Optimization

Randomizing strategies are used in two-person games.

Consider game scissors-stone-paper, for which the optimal strategy is to randomize with probabilities $1/3$, $1/3$, $1/3$.

In the Optimization course you will learn how to solve other games.

Here is one you will be able to solve (from a recent Ph.D. thesis):

> I have lost k possessions in my room (keys, wallet, phone, etc). Searching location i costs c_i.
> "Sod's Law" predicts that I will have lost my objects in whatever way makes finding them most difficult.
>
> Assume there are n locations, and cost of searching location i is c_i. I will search until I find all my objects.
>
> What is Sod's Law? How do I minimize my expected total search cost? (I will have to randomize.)

Index

1 Classical probability

Classical probability. Sample spaces. Equally likely outcomes. *Equalizations of heads and tails*. *Arcsine law*.

1.1 Diverse notions of 'probability'

Consider some uses of the word 'probability'.

1. The probability that a fair coin will land heads is $1/2$.
2. The probability that a selection of 6 numbers wins the National Lottery Lotto jackpot is 1 in $\binom{49}{6}$ =13,983,816, or 7.15112×10^{-8}.
3. The probability that a drawing pin will land 'point up' is 0.62.
4. The probability that a large earthquake will occur on the San Andreas Fault in the next 30 years is about 21%.
5. The probability that humanity will be extinct by 2100 is about 50%.

Clearly, these are quite different notions of probability (known as classical[1,2], frequentist[3] and subjective[4,5] probability).

Tripos questions 2015

The end