37262 Mathematical Statistics

Lecture 5

UTS CRICOS 00099F

How Predictable are Random Samples?

- Consider the problem of predicting what value a sample mean might take, given a sample size and the distribution from which independent realisations are drawn.
- Assuming that the variance of the distribution is non-zero, we cannot ever know for certain what value a sample mean can take.
- We can, though, put some bounds on which ranges of possible values are more or less probable.
- Intuitively, sample means from larger samples should be more predictable, as should samples from distributions with lower variance.

How Predictable are Random Samples?

- Consider now a non-negative continuous random variable X with density function $f(x)$.

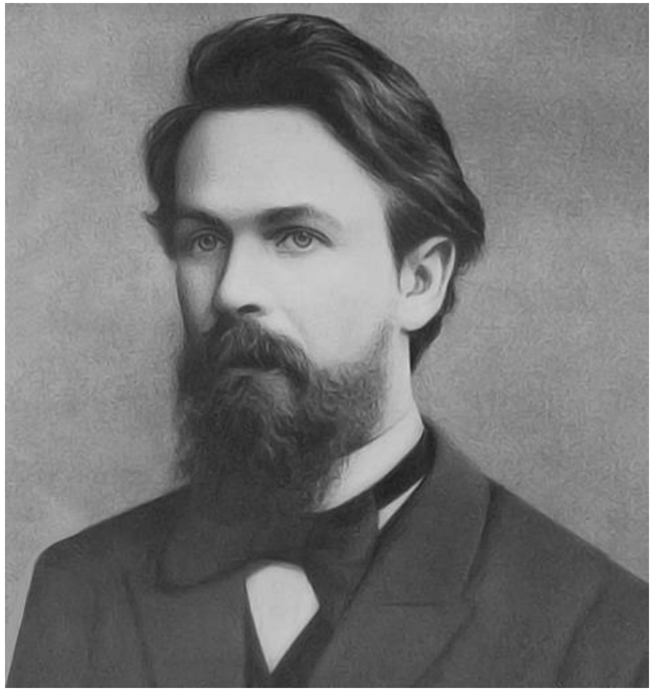
- For any $a > 0$, we have that $E(X) = \int_0^\infty xf(x)dx = \int_0^a xf(x)dx + \int_a^\infty xf(x)dx$

  $\geq \int_a^\infty xf(x)dx \geq \int_a^\infty af(x)dx$ since $x \in [a,\infty)$ implies $x \geq a$.

- We therefore have that $E(X) \geq a\int_a^\infty f(x)dx$

Markov’s Inequality

- $E(X) \geq a\int_a^{\infty} f(x)dx$ implies that, for $a > 0$, we have $\frac{E(X)}{a} \geq P(X \geq a)$.
- This is **Markov’s Inequality**.
- This result also holds for discrete random variables. The proof is essentially the same, with sums rather than integrals.



Andrey Markov

(*1856 – 1922*)

A Related Result

- Let X be any random variable such that $Var(X) > 0$.

- Consider now the random variable defined as $Y = \frac{(X - E(X))^2}{Var(X)}$.

- By definition, we have that $E(Y) = E\left[\frac{(X - E(X))^2}{Var(X)}\right] = \frac{1}{Var(X)} E(X - E(X))^2 = 1$

- Applying Markov's Inequality to the variable Y (which is non-negative valued, since it is the ratio of a squared value to a variance, neither of which can be negative) gives

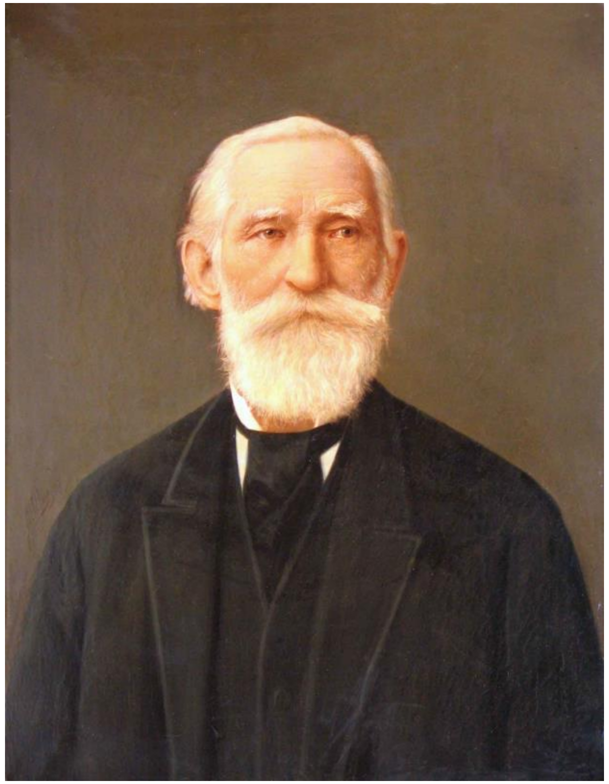
  $P(Y \geq k^2) \leq \frac{E(Y)}{k^2}$ so $P\left(\frac{(X - E(X))^2}{Var(X)} \geq k^2\right) \leq \frac{1}{k^2}$ for any real value k.

Chebyshev's Inequality

- $P\left(\frac{(X-E(X))^2}{Var(X)} \geq k^2\right) \leq \frac{1}{k^2}$.
- Noting that $P\left(\frac{(X-E(X))^2}{Var(X)} \geq k^2\right) = P\left(|X-E(X)| \geq k\sqrt{Var(X)}\right)$

  we get that $P\left(|X-E(X)| \geq k\sqrt{Var(X)}\right) \leq \frac{1}{k^2}$.

- This is **Chebyshev's Inequality**.



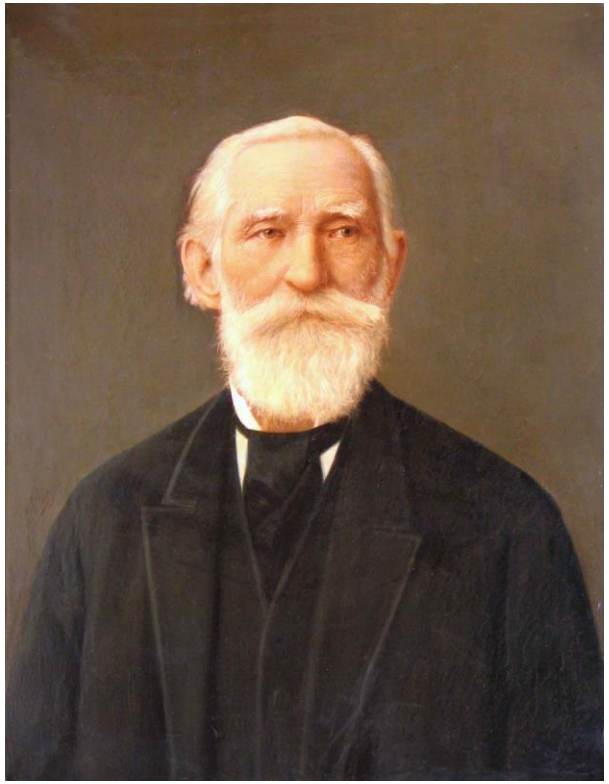
Pafnuty Chebyshev

(*1821– 1894*)

Chebyshev's Inequality

- Chebyshev's Inequality $P\left(|X-E(X)|\geq k\sqrt{Var(X)}\right)\leq\frac{1}{k^2}$ tells us that, for variables drawn from any distribution, no more than $\frac{1}{k^2}$ of observations can be expected to lie more than *k* standard deviations above or below the mean.

- For example, no more than 25% of observations can be expected be more than two standard deviations away from the mean and no more than 6.25% of observations can be expected to be more than four standard deviations from the mean.



Pafnuty Chebyshev

(*1821– 1894*)

Cantelli's Inequality

- There is a similar inequality to Chebyshev's Inequality but only for one-sided differences between the a random variable and its expectation.

- Let X be any random variable such that $Var(X) = \sigma^2 > 0$.

- Consider now the random variable $Q = X - E(X)$. Clearly $E(Q) = 0$ and $Var(Q) = \sigma^2$.

- For any $q, t > 0$, $P(Q \geq q) = P(Q + t \geq q + t)$

Cantelli's Inequality

- $P\left[(Q+t)^2 \geq (q+t)^2\right] \leq P[(Q+t) \geq (q+t)]$, since $[(Q+t) \geq (q+t)]$ implies that $\left[(Q+t)^2 \geq (q+t)^2\right]$ but $\left[(Q+t)^2 \geq (q+t)^2\right]$ does not imply that $[(Q+t) \geq (q+t)]$.

- This gives $P(Q \geq q) \leq P\left[(Q+t)^2 \geq (q+t)^2\right]$.

- Applying Markov's Inequality to the right-hand side, we have that $P(Q \geq q) \leq \dfrac{E\left[(Q+t)^2\right]}{(q+t)^2}$

- $P(Q \geq q) \leq \dfrac{\sigma^2 + t^2}{(q+t)^2}$. We can establish a tighter bound by minimising as a function of t.

Cantelli's Inequality

- $$\frac{\partial}{\partial t}\left[\frac{\sigma^2+t^2}{(q+t)^2}\right]=\frac{2t}{(q+t)^2}-2\frac{(\sigma^2+t^2)}{(q+t)^3}=\frac{2qt-2\sigma^2}{(q+t)^3}$$

- This is minimised when $\frac{\partial}{\partial t}\left[\frac{\sigma^2+t^2}{(q+t)^2}\right]=0$ i.e. when $t=\frac{\sigma^2}{q}$.

- We then obtain $P(Q\geq q)\leq\frac{\sigma^2+\left(\frac{\sigma^2}{q}\right)^2}{\left(q+\left(\frac{\sigma^2}{q}\right)\right)^2}=\frac{q^2\sigma^2+\sigma^4}{\left(\frac{q^2+\sigma^2}{q}\right)^2}=\frac{\sigma^2}{q^2+\sigma^2}$

- This therefore gives that $P(X-E(X)\geq q)\leq\frac{\sigma^2}{q^2+\sigma^2}$ where $Var(X)=\sigma^2$.

Cantelli's Inequality

- This is **Cantelli's Inequality**.

- One consequence of this is that, setting $q = \sqrt{\sigma^2}$ into $P(X - E(X) \geq q) \leq \frac{\sigma^2}{q^2 + \sigma^2}$, we obtain that $P(X - \mu \geq \sigma) \leq \frac{1}{2}$ where $E(X) = \mu$.

- $P(X \geq \sigma + \mu) \leq \frac{1}{2}$.

Francesco Cantelli

(*1875 – 1966*)

Cantelli's Inequality

- Applying the same argument to the variable $-X$ which, by definition has $E(-X) = -\mu$ and $Var(-X) = \sigma^2$, we also obtain $P(-X \geq \sigma - \mu) \leq \frac{1}{2}$ which implies that $P(X \leq \mu - \sigma) \leq \frac{1}{2}$.

- The median, X_{med}, of the distribution of X, by definition, satisfies $P(X \geq X_{med}) \geq \frac{1}{2}$ and $P(X \leq X_{med}) \geq \frac{1}{2}$.

- Together, these give us that $|X_{med} - \mu| \leq \sigma$ i.e. for any distribution, the mean and median cannot differ by more than one standard deviation.

Francesco Cantelli

(*1875 – 1966*)

A Note on Inequalities

- It should be noted that Markov's, Chebyshev's and Cantelli's Inequalities provide bounds on probabilities and not estimates of them.

- For example, with a standard normal variable, $Z \sim N(0,1)$. Chebyshev's Inequality tells us that at least 75% of realisations of the variable are expected to lie no more than two standard deviations from the mean. In fact

$$\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \approx 0.9545 > 0.75 \quad \text{and} \quad \int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \approx 0.9973 > \frac{8}{9}$$

- Although we know that for any distribution, the mean and median cannot differ by more than one standard deviation, for $Z \sim N(0,1)$, they do not differ at all – both are zero.

A Related Result

- Consider now a sequence of independent identically distributed variables $X_1, X_2, \ldots, X_n$ each with $E(X_i) = \mu < \infty$ and $Var(X_i) = \sigma^2 < \infty$.

- We therefore have that $E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n}E(X_1 + X_2 + \ldots + X_n) = \frac{1}{n}(\mu + \mu + \ldots + \mu) = \mu.$

- $Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n^2}Var(X_1 + X_2 + \ldots + X_n) = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2) = \frac{\sigma^2}{n}$
  (since the variables are independent, hence uncorrelated.)

- Applying Chebyshev's Inequality, we get $P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$

A Related Result

- Setting $\varepsilon = \frac{k\sigma}{\sqrt{n}}$ into $P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$ gives

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

- Now, as the number of variables $n \to \infty$ then, for any fixed $\mu < \infty, \sigma^2 < \infty$,

$$\lim_{n\to\infty}\left[P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right)\right] = 0.$$

- In other words, as the sample size gets extremely large, the probability that the average of the variables is anything other than μ converges to zero.

Weak Law of Large Numbers

- This is the **Weak Law of Large Numbers**, $\lim_{n\to\infty}\left[\frac{X_1+X_2+...+X_n}{n}\right]=\mu$ with probability 1

  where each X_i is independent and identically distributed with finite expectation μ and finite variance.

- The long-run average of any sequence of random variables with finite expectation and finite variance converges to the expected value of those variables with probability 1.

Monte Carlo Integration

- Consider now the problem of evaluating the definite integral $\int_0^1 g(x)dx$.
- Let $U_1, U_2, ..., U_n$ be a sequence of independent variables such that each $U_i \sim U[0,1]$.
- Now, since each $U_i \sim U[0,1]$, the density function of each is $f(u_i) = \begin{cases} 1 & u_i \in [0,1] \\ 0 & \text{otherwise} \end{cases}$.

- This therefore gives $\int_0^1 g(x)dx = \int_0^1 [1 \times g(x)]dx = \int_0^1 [1 \times g(u_i)]du_i = E(g(U_i))$.
- This tells us that, for $U_i \sim U[0,1]$, then $g(U_i)$ is a random variable (with unknown distribution) with expectation $E(g(U_i)) = \int_0^1 g(x)dx$.

Monte Carlo Integration

- As we know that $g(U_1), g(U_2), \ldots, g(U_n)$ is a sequence of independent random variables such that $E(g(U_i)) = \int_0^1 g(x)dx$, we can apply the Weak Law of Large Numbers to obtain that

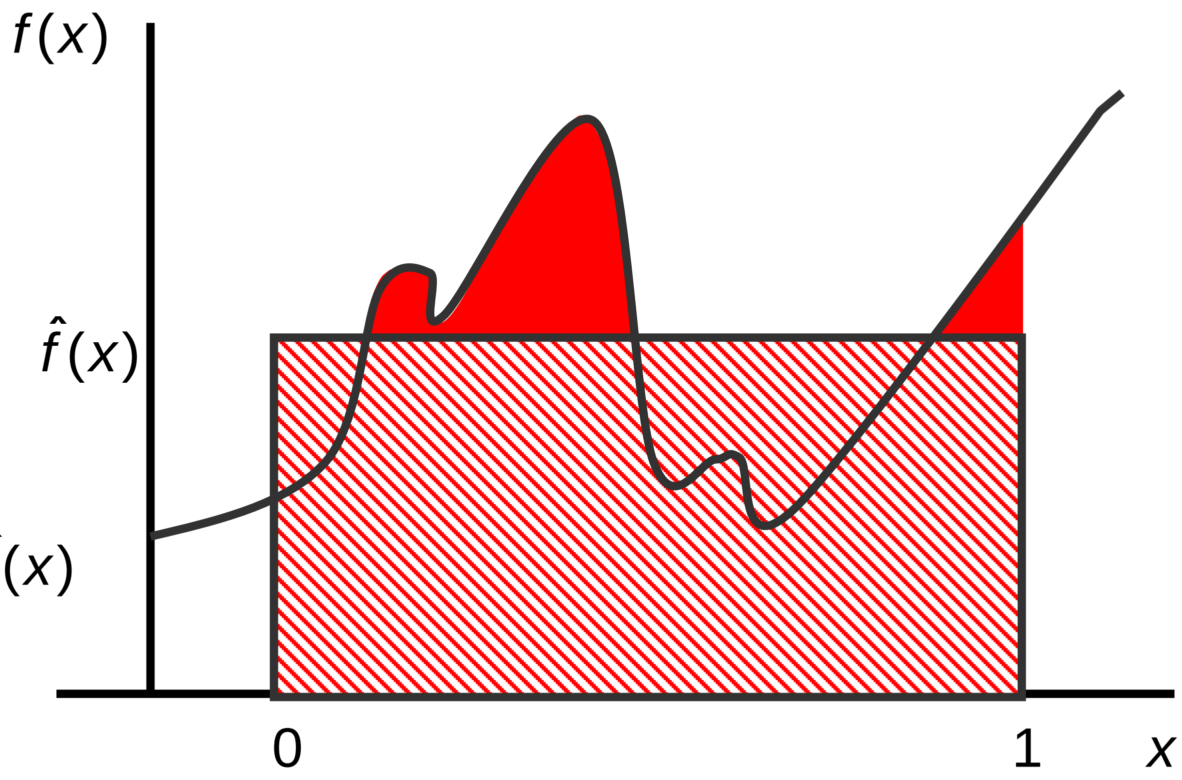
$$\frac{1}{n}\sum_{i=1}^{n} g(U_i) \approx \int_0^1 g(x)dx.$$

- We know that, as $n \to \infty$, the difference between the two sides in the approximation above tends to zero with probability 1.

Monte Carlo Integration

- Consider evaluating the definite integral of a function $f(x)$ between 0 and 1.

- Assuming the associated area is finite, the value of the integral is equal to that of a rectangle with length 1 and height $\hat{f}(x)$ where $\hat{f}(x)$ is the average value of the function over that region.



- An estimate of this area can be found simply by taking a large number of independent $U[0,1]$ variables, $u_1, u_2, \ldots, u_n$, and averaging to find $\hat{f}(x) \approx \dfrac{f(u_1) + f(u_2) + \ldots + f(u_n)}{n}$

Monte Carlo Integration: Example

$$\frac{1}{1+0.935146^2} \approx 0.533476$$

- Consider now estimating the value of $\int_0^1 \frac{1}{1+x^2}\,dx$ by the Monte Carlo method.
- Columns A – E each contain 20 realisations of a $U[0,1]$ variable.
- Columns G – K each contain 20 realisations of $\frac{1}{1+u_i^2}$ for the corresponding value in columns A – E.

$$\frac{1}{1+0.164711^2} \approx 0.973587$$

	A	B	C	D	E	F	G	H	I	J	K
1	0.935146	[illegible]	0.439073	0.839565	0.314174		0.533476	0.883772	0.838374	0.586555	0.910162
2	0.123136	0.571505	0.810423	0.417027	0.137596		[illegible]	0.753797	0.603578	0.851853	0.981419
3	0.382501	0.347116	0.364187	0.135789	0.871119		0.872366	0.892467	0.882899	0.981895	0.568554
4	0.300543	0.720004	0.415703	0.78995	0.724183		0.917157	0.658585	0.852654	0.615756	0.655978
5	0.328152	0.241302	0.180209	0.06861	0.19947		0.902785	0.944977	0.968546	0.995315	0.961734
6	0.961071	0.784047	0.707974	0.598807	0.097218		0.519843	0.619299	0.666122	0.736068	0.990637
7	0.999031	0.938459	0.734723	0.437605	0.805962		0.500485	0.531716	0.649428	0.839279	0.606217
8	0.516444	0.311315	0.535394	0.616934	0.638913		0.789444	0.911646	0.777214	0.724319	0.710121
9	0.763581	0.001963	0.610188	0.174166	0.764827		0.631689	0.999996	0.728688	0.970559	0.63093
10	0.2831	0.839002	0.649126	0.248825	0.423268		0.925801	0.58688	0.703549	0.941696	0.848065
11	0.905064	0.77959	0.91391	0.621309	0.015811		0.54971	0.621983	0.544891	0.721488	0.99975
12	0.899732	0.266889	0.168596	0.49815	0.718692		0.552633	0.933506	0.972361	0.801184	0.659405
13	0.808893	0.854261	0.576185	0.963141	0.809653		0.604483	0.578114	0.750757	0.518769	0.604033
14	0.734345	0.367821	0.552804	0.008556	0.936555		0.649662	0.88083	0.765936	0.999927	0.532727
15	0.474311	0.125894	[illegible]	0.646228	0.528801		0.816346	0.984398	[illegible]	0.705412	0.781475
16	0.208938	[illegible]	0.164711	0.28725	0.878091		0.958171	[illegible]	0.973587	0.923777	0.564639
17	0.835166	0.789315	0.153141	0.72874	0.362848		0.589101	0.616134	0.977085	0.653142	0.883659
18	0.273949	0.816553	0.187365	0.310286	[illegible]		0.930191	0.599967	0.966085	0.912178	0.974911
19	0.669444	0.402979	0.201919	0.225738	0.709133		0.690534	0.860295	0.960826	0.951513	0.665393
20	0.303085	0.260821	0.140529	0.191407	0.099072		0.915868	0.936305	0.980634	0.964658	0.99028

Monte Carlo Integration: Example

- Averaging the first 20 variables gives

$$\int_0^1 \frac{1}{1+x^2}dx \approx \frac{\frac{1}{1+u_1^{\,2}} + \frac{1}{1+u_2^{\,2}} + \ldots + \frac{1}{1+u_{20}^{\,2}}}{20}$$

$$\approx 0.7417$$

- The estimates after 40, 60, 80 and 100 are 0.7570, 0.7773 and 0.7879 respectively.

- The exact value of $\int_0^1 \frac{1}{1+x^2}dx$

is $\arctan(1) - \arctan(0) = \frac{\pi}{4} \approx 0.7854.$

	A	B	C	D	E	F	G	H	I	J	K
1	0.935146	0.362648	0.439073	0.839565	0.314174		0.533476	0.883772	0.838374	0.586555	0.910162
2	0.123136	0.571505	0.810423	0.417027	0.137596		0.985064	0.753797	0.603578	0.851853	0.981419
3	0.382501	0.347116	0.364187	0.135789	0.871119		0.872366	0.892467	0.882899	0.981895	0.568554
4	0.300543	0.720004	0.415703	0.78995	0.724183		0.917157	0.658585	0.852654	0.615756	0.655978
5	0.328152	0.241302	0.180209	0.06861	0.19947		0.902785	0.944977	0.968546	0.995315	0.961734
6	0.961071	0.784047	0.707974	0.598807	0.097218		0.519843	0.619299	0.666122	0.736068	0.990637
7	0.999031	0.938459	0.734723	0.437605	0.805962		0.500485	0.531716	0.649428	0.839279	0.606217
8	0.516444	0.311315	0.535394	0.616934	0.638913		0.789444	0.911646	0.777214	0.724319	0.710121
9	0.763581	0.001963	0.610188	0.174166	0.764827		0.631689	0.999996	0.728688	0.970559	0.63093
10	0.2831	0.839002	0.649126	0.248825	0.423268		0.925801	0.58688	0.703549	0.941696	0.848065
11	0.905064	0.77959	0.91391	0.621309	0.015811		0.54971	0.621983	0.544891	0.721488	0.99975
12	0.899732	0.266889	0.168596	0.49815	0.718692		0.552633	0.933506	0.972361	0.801184	0.659405
13	0.808893	0.854261	0.576185	0.963141	0.809653		0.604483	0.578114	0.750757	0.518769	0.604033
14	0.734345	0.367821	0.552804	0.008556	0.936555		0.649662	0.88083	0.765936	0.999927	0.532727
15	0.474311	0.125894	0.506881	0.646228	0.528801		0.816346	0.984398	0.795591	0.705412	0.781475
16	0.208938	0.73416	0.164711	0.28725	0.878091		0.958171	0.649776	0.973587	0.923777	0.564639
17	0.835166	0.789319	0.153141	0.72874	0.362848		0.589101	0.616134	0.977085	0.653142	0.883659
18	0.273949	0.816553	0.187365	0.310286	0.160421		0.930191	0.599967	0.966085	0.912178	0.974911
19	0.669444	0.402979	0.201919	0.225738	0.709133		0.690534	0.860295	0.960826	0.951513	0.665393
20	0.303085	0.260821	0.140529	0.191407	0.099072		0.915868	0.936305	0.980634	0.964658	0.99028

Monte Carlo Integration

- As originally formulated, the Monte Carlo method approximates only definite integrals evaluated between 0 and 1.

- We can, though, change the limits of a definite integral though an appropriate substitution.

- For example, to apply the Monte Carlo method to approximate the value of $\int_7^{10} g(x)dx$, we first need to change the domain of integration to between 0 and 1.

- An appropriate substitution is therefore $y = \frac{x-7}{3}$ since when $x = 7, y = 0$ and when $x = 10, y = 1$.

- This gives $\int_7^{10} g(x)dx = \int_0^1 g(3y+7)\frac{dx}{dy}dy = \int_0^1 3g(3y+7)dy \approx \frac{1}{n}\sum_{i=1}^{n} 3g(3u_i + 7)$

Monte Carlo Integration

- Similarly, definite integrals across infinite domains can be evaluated with appropriate substitutions.
- Consider the problem of evaluating $\int_{-\infty}^{\infty} e^{-x^2} dx$.
- We require a change of variables such that we have an integral with respect to $y \in [0,1]$ such that $\lim_{x \to -\infty} y(x) = 0$ and $\lim_{x \to \infty} y(x) = 1$.
- There are multiple choices for this, for example $y = \frac{e^x}{1+e^x}$ maps the whole real line $(-\infty, \infty)$ to $[0,1]$, but so do $y = \frac{e^{2x}}{1+e^{2x}}, y = \frac{e^{3x}}{1+e^{3x}}$ etc.
- There are "better" choices of substitution to give quicker convergence, but this is a more advanced topic

Monte Carlo Integration: Example

- Here, we will approximate the value of $\int_{-\infty}^{\infty} e^{-x^2} dx$ by substituting $y = \frac{e^x}{1+e^x}$.
- This gives $y = \frac{e^x}{1+e^x}$ so $y + ye^x = e^x$ or $e^x = \left(\frac{y}{1-y}\right)$.
- $x = \ln\left(\frac{y}{1-y}\right)$.
- Also, $\frac{dy}{dx} = \frac{e^x}{(1+e^x)^2} = y(1-y)$.
- Substituting in gives $\int_{-\infty}^{\infty} e^{-x^2} dx = \int_0^1 \frac{e^{-\left[\ln\left(\frac{y}{1-y}\right)\right]^2}}{y(1-y)} dy$
- We can therefore obtain an estimate of this by averaging many realisations of $\frac{e^{-\left[\ln\left(\frac{u_i}{1-u_i}\right)\right]^2}}{u_i(1-u_i)}$ where each $U_i \sim U[0,1]$.

Monte Carlo Integration: Example

- Averaging 100,000 realisations, the mean of these gave $\int_{-\infty}^{\infty} e^{-x^2}dx \approx 1.7729$.
- We compare with the normal distribution function $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$, hence our initial integral is that of a normal density, with $\mu = 0, \sigma^2 = 0.5$.
- The exact value of the integral is therefore $\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi} \approx 1.7725$.

Possible Changes of Variable

- Below is a table of suggested substitutions for each of the four cases you may encounter:

x domain	$y(x)$	$x(y)$	$\frac{dy}{dx}$
$x \in [a,b]$	$y = \frac{x-a}{b-a}$	$x = a + (b-a)y$	$\frac{dy}{dx} = \frac{1}{b-a}$
$x \in [a,\infty)$	$y = 1 - e^{a-x}$	$x = a - \ln(1-y)$	$\frac{dy}{dx} = e^{a-x} = 1-y$
$x \in (-\infty,b]$	$y = e^{x-b}$	$x = b + \ln(y)$	$\frac{dy}{dx} = e^{x-b} = y$
$x \in (-\infty,\infty)$	$y = \frac{e^x}{1+e^x}$	$x = \ln\left(\frac{y}{1-y}\right)$	$\frac{dy}{dx} = \frac{e^x}{(1+e^x)^2} = y(1-y)$