Lösung zur Monte-Carlo Integration

- Berechnen Sie
$$\int_0^1 \int_0^1 \int_0^1 \exp(x+y+z) dx dy dz$$
mittels einer geeigneten Monte-Carlo Simulation!

- Erste Beobachtung: $e^0 \leq e^{x+y+z} \leq e^3$ für $(x, y, z) \in [0, 1]^3$.

- Hier zwei Lösungen, für die beiden prinzipiellen Herangehensweisen nutzen.

Lösung mit *hit-or-miss*

```
lower.limit <- 0
upper.limit <- exp(3)
runs <- 1000
hitcounter <-0
for (run in 1:runs){
   measurement.point <- runif(3)
   point.value <- exp(sum(measurement.point))
   hitormiss <- runif(1, lower.limit, upper.limit)
   if (hitormiss <= point.value) hitcounter <- hitcounter + 1
}
upper.limit*hitcounter/runs
[1] 5.121812
```

Lösung über den Erwartungswert einer Funktion

```
runs <- 1000
integralsum<-0
for (run in 1:runs){
   measurement.point <- runif(3)
   point.value <- exp(sum(measurement.point))
   integralsum <- integralsum + point.value
}
integralsum/runs
[1] 5.040594
```

Genauigkeitsüberlegungen

- Da durch die Monte-Carlo Verfahren die exakten numerischen Lösungen mathematischer Aufgaben durch Zufallsvariablen (i.d.R. Mittelwerte vieler Zufallsvariablen) ersetzt wurden, bleibt die Frage, wie man die Genauigkeit der Lösungen angeben kann.

- Gesucht ist ein Maß, welches erlaubt, analog zur numerischen Mathematik anzugeben, wie viele der angegebenen Stellen "signifikant" sind.

- Bei Zufallsvariablen wird diese Genauigkeit mittels der aus Statistik II bekannten Konfidenzintervalle zu einem gewünschten Niveau $1 - \alpha$ angegeben.

- Da bei Monte-Carlo-Simulationen große Stichprobenumfänge kein Problem sind, kommen uns das Gesetz der großen Zahlen und der zentrale Grenzwertsatz (Statistik II) zur Hilfe.

- Zur Erinnerung: Für i.i.d. ZV X_i, $i = 1, \ldots, n$ mit $E(X_i) = \mu < \infty$ und $\text{Var}(X_i) = \sigma^2 < \infty$ sagt der zentrale Grenzwertsatz

$$\lim_{n \to \infty} \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1).$$

- Damit ergibt sich das $100(1 - \alpha)$%-Konfidenzintervall für μ zu:

$$[\bar{X} - \Phi^{-1}(1 - \alpha/2) \frac{S}{\sqrt{n}}; \bar{X} + \Phi^{-1}(1 - \alpha/2) \frac{S}{\sqrt{n}}],$$

mit unbekanntem $\sigma^2, n > 50, S^2 := \frac{1}{n-1} \sum_1^n (X_i - \bar{X})^2$.

- Der Stichprobenumfang (Anzahl der Replikationen der Simulation) bestimmt die länge des KI über den Faktor $\frac{1}{\sqrt{n}}$. Mit der Wurzel aus der Anzahl der Replikationen geht die Breite des Konvergenzintervalls gegen 0.

- Im Umkehrschluss bedeutet dies, dass für eine zusätzliche Dezimalstelle der Stichprobenumfang verhundertfacht werden muss!

- Für festes α ist bei $n = 100$ Wiederholungen das Konfidenzintervall proportional zu $0.1S$!

- Die andere Stellschraube, um das Konfidenzintervall zu verkürzen ist der Varianzschätzer S^2. Jede Änderung der Standardabweichung S geht linear in die Länge des Intervalls ein.

- Oder man ändert α. Aber das wird in der Regel nicht gewünscht sein.

Genauigkeit am Beispiel

- Das Integral aus der Übungsaufgabe lässt sich trivial direkt lösen:

$$I = \int_0^1 \int_0^1 \int_0^1 \exp(x+y+z)\text{dx dy dz} = (e^1 - e^0)^3 \approx 5.073214.$$

- Die beiden Ansätze werden zunächst getrennt betrachtet und dann verglichen.

Überlegungen für den Ansatz *hit or miss*

- Eine Simulationsdurchlauf erzeugt eine Zufallsvariable X gemäß einer Bernoulli-Verteilung Ber(p). Der Wert der ZV X ist Vol(Q), wenn das gesuchte Gebiet bei hit-or-miss getroffen wird, 0 sonst.

- Daraus folgt sofort: $p = I/\text{Vol}(Q)$, wobei Vol(Q) das Volumen des gewählten umschliessenden Quaders Q ist. Im Fall der obigen Lösung gilt $Vol(Q) = 1 \cdot 1 \cdot 1 \cdot e^3 = e^3 = 20.08554$.

- Es gelten dann:

$$E(X) = p\text{Vol}(Q) = I/\text{Vol}(Q) \cdot \text{Vol}(Q) = I$$

und
$\mathsf{Var}(X) = p\mathsf{Vol}(Q)^2 - p^2\mathsf{Vol}(Q)^2 = I\mathsf{Vol}(Q) - I^2 = I(\mathsf{Vol}(Q) - I)$.

- Man sieht also, dass die Varianz einer Simulation direkt proportional zum Volumen des umschliessenden Quaders ist.

- Setzt man die entsprechende Größen für die Übungsaufgabe ein, so gilt: $E(X) = 5.0732$, $\sqrt{Var(X)} \approx \sqrt{5.07(20.08 - 5.07)} = 8.73$.

- Ganz allgemein gilt nach dem zentralen Grenzwertsatz bekanntlich für große n, dass

$$\bar{X} = \frac{1}{n}\sum_{1}^{n} X_i \sim N\left(E(X), \frac{\mathsf{Var}(X)}{n}\right).$$

- Für die Breite $|KI_\alpha|$ des Konfidenzintervalls zum Niveau $1-\alpha$ gilt genauso bekanntlich

$$|KI_\alpha| = 2 * \Phi^{-1}(1 - \frac{\alpha}{2})\frac{\sigma}{\sqrt{n}}$$

für gegebenes α und n. Die folgende Tabelle gibt die Breite dieser Intervalle für verschiedene Wahlen von α und n für die Situation der Übungsaufgabe an.

$1-\alpha \backslash n$	100	1000	10000	250000	1e6
0.9	2.870	0.907	0.287	0.057	0.028
0.95	3.420	1.081	0.342	0.068	0.034
0.975	3.912	1.237	0.391	0.078	0.039
0.99	4.495	1.421	0.449	0.089	0.044
0.999	5.743	1.816	0.574	0.114	0.057

- Analog lässt sich zu einer vorgegebenen Intervallbreite $|KI_\alpha|$ ganz allgemein der mindestens benötigte Stichprobenumfang n herleiten, um diese Länge zu unterschreiten:

$$n \geq \left(\frac{2\Phi^{-1}(1-\frac{\alpha}{2})\sigma}{|KI_\alpha|} \right)^2 .$$

Überlegungen für die Simulation des Integrals als Erwartungswert einer Funktion $f(X), X \sim U_{[0,1]}$

- Auch hier lassen sich noch E-Wert und Varianz direkt berechnen.
- Der Erwartungswert ist wie oben: $E(f(X)) = 5.0732$, die Varianz ergibt sich zu

$$\begin{aligned} Var(X) &= \int_{[0,1]^3} (e^{x+y+z})^2 \mathsf{dx\ dy\ dz} - E(X)^2 \\ &= \int_0^1 \int_0^1 \int_0^1 e^{2x} \mathsf{dx} e^{2y} \mathsf{dy} e^{2z} \mathsf{dz} - (e-1)^6 \\ &= \frac{1}{8}(e^2-1)^3 - (e-1)^6 \approx 6.862688. \end{aligned}$$

- Und folglich $\sqrt{Var(X)} \approx 2.62$. Da die Standardabweichung linear in die Inervalllängen eingeht, sind die KIe mit dieser Methode für die Übungsaufgabe um ca. $2/3 (= 1 - 2.62/8.73)$ kürzer!

- Damit ergeben sich die Konfidenzintervalllängen $|KI_\alpha|$ zu:

$1-\alpha \backslash n$	100	1000	10000	250000	1e6
0.9	0.861	0.272	0.086	0.017	0.008
0.95	1.026	0.324	0.102	0.020	0.010
0.975	1.174	0.371	0.117	0.023	0.011
0.99	1.349	0.426	0.134	0.026	0.013
0.999	1.724	0.545	0.172	0.034	0.017

Aufgabe: Konfidenzintervalle für Monte-Carlo Schätzer

- Normalerweise sind weder exakter Erwartungswert noch die exakte Varianz berechenbar, dann werden jeweils die bekannten Schätzer aus Statistik II für die Bestimmung des Konfidenzintervalls genutzt.

- Welcher Stichprobenumfang n wäre bei den beiden Herangehensweisen jeweils nötig, um eine Konfidenzintervalllänge von 0.1 bei $\alpha = 0.05$ für das Beispiel der Übung zu erreichen?

- Schreiben Sie die Simulationen für die Monte-Carlo-Integration für den 2. Fall so in eine Funktion um, dass als Argumente die Anzahl der Einzelsimulationen und ein Niveau α angegeben werden können. Als Ergebnis soll der Mittelwertschätzer und die Breite des Konfidenzintervalls bei dem angegebenen Niveau α ausgegeben werden.

Lösung:

```
mcfunk <- function(runs=1000, alpha=0.05)
  {
    point.values <- rep(0, runs)
    for (i in 1:runs){
      measurement.point <- runif(3)
      point.values[i] <- exp(sum(measurement.point))
    }
    integralsum <- mean(point.values)
    estsd <- sqrt(var(point.values))
    cat("Bei ", runs, " Wiederholungen ergibt sich ein
                          1-",alpha,"KI von [")
    cat(integralsum - qnorm(1-alpha/2)*estsd/sqrt(runs), " ; ")
    cat(integralsum + qnorm(1-alpha/2)*estsd/sqrt(runs), "]\n")
  }
```

Varianzreduzierende Techniken

- Eine Erhöhung des Stichprobenumfangs ist zwar möglich, wegen der langsamen Konvergenzgeschwindigkeit jedoch nur bedingt nützlich.

- Eine Verfahrensverbesserung, die zu einer Reduktion der empirischen Varianz S^2 führt, muss dagegen nur einmal hergeleitet werden und reduziert die erforderlichen Stichprobenumfänge für alle folgenden Simulationen!

- Diesem Ansatz wurde und wird deshalb erheblicher Forschungsaufwand gewidmet.

- Im Folgenden werden einige kurze Bemerkungen zu den gebräuchlichsten Möglichkeiten der Varianzreduktion für Simulationsstudien gegeben.

Reduktion der Varianz bei *hit-or-miss*

- Die Varianz der ZV X, auf der diese Methode beruht, war gegeben durch $\mathrm{Var}(X) = I(\mathrm{Vol}(Q) - I))$.

- Die Varianz kann also direkt durch die Verkleinerung des umschliessenden Quaders reduziert werden.

- Es lohnt sich also, die zu integrierende Funktion genau anzuschauen.

- In unserem Beispiel kann man z.B. sehen, dass gilt:

$$e^{x+y+z} \geq 1 \text{ für alle } (x, y, z) \in [0, 1]^3.$$

  Es würde also eine Varianzreduktion erreicht, wenn man die Zufallszahl für den *hit-or-miss* Versuch nicht aus $[0, e^3]$ wählt, sondern aus $[1, e^3]$!

- Natürlich muss der Schätzer entsprechend korrigiert werden, da ja der Bereich $[0,1]^4$ zum gesuchten Integral gehört!

- Es ist weiterhin zu bemerken, dass die Methode nicht essentiell darauf beruht, dass *ein* Quader betrachtet betrachtet wird. Je näher man mit einer umschliessenden Funktion, deren Integral bekannt ist, dem Kurvenverlauf der zu integrierenden Funktion kommt, desto kleiner wird die Varianz des Schätzers! Man könnte also die Definitionsmenge geschickt aufteilen und jeden einzelnen Teil mit einem eigenen Quader, der möglichst eng gefasst ist approximieren.

- Schließlich ist es auch nicht entscheidend, dass Quader gewählt werden. Wichtig ist nur, dass man in der gewählten Form gleichverteilte ZV generieren kann.

Aufgabe: Varianzreduktion bei *hit-or-miss*:

- Modifizieren Sie die Lösung der Übungsaufgabe entsprechend der Bemerkung über den kleineren Quader!
- Berechnen sie die Varianz dieses Verfahrens! Was bedeutet dies für die Intervalllängen im Vergleich zur ersten Lösung?
- Überprüfen Sie diese Berechnungen mittels einer Simulation!

Varianzreduktion durch antithethische Variblen

- Eine erstaunliche und erstaunlich gut funktionierende Idee zur Varianzreduktion besteht in der Nutzung sogenannter *antithetischer (Zufalls-)Variablen*.

- Statt jeweils nur eine Zufallszahl X aus $U_{[0,1]}$ zu nutzen, nutzt man zu jedem solchen X gleichzeitig auch $Y := 1 - X$. Es gilt $Y \sim U_{[0,1]}$.

- Zunächst mal kann beobachtet werden, dass nur die halbe Anzahl von Zufallszahlen erzeugt werden muss um eine bestimmte Anzahl von Zufallszahlen zu erhalten.

- Sei nun eine Stichprobe $X_1, \ldots, X_n \sim U_{[0,1]}$ i.i.d. gegeben. Dann ist auch $1 - X_1 =: Y_1, \ldots, 1 - X_n =: Y_n \sim U_{[0,1]}$ i.i.d. eine solche Stichprobe.

- Ist nun $\hat{\theta}_1 = f(X_1, \ldots, X_n)$ ein e-treuer Schätzer für θ, so ist dies auch $f(Y_1, \ldots, Y_n) = \hat{\theta}_2$.

- Damit gilt $E(\frac{\hat{\theta}_1 + \hat{\theta}_2}{2}) = \theta$.

- Man kann zeigen, dass antithetische Variablen zu einer Varianzreduktion führen, wenn die Schätzer $\hat{\theta}_1$ und $\hat{\theta}_2$ negativ korreliert sind.

- Dieselbe Idee kann auch genutzt werden, falls die Inversionsmethode zur Zufallszahlenerzeugung genutzt wird. Ein antithetisches Paar von Zufallszahlen ist dann durch $F^{-1}(U)$ und $F^{-1}(1 - U)$ gegeben.

- **Aufgabe:** Probieren Sie diese Methode für das Beispiel der Übung zu heute aus!

Varianzreduktion mittels *Importance Sampling*

- *Importance Sampling* ist eine Verbesserung der einfachen Methode ein Integral als Erwartungswert zu approximieren.

- Zur Erinnerung: Im naiven Ansatz wählt man Zufallszahlen x_i gleichverteilt im Definitionsbereich einer Funktion f. Dann schätzt man das Integral $I = \int_D f(x)\text{dx}$ durch

$$\hat{I} = \frac{1}{n} \sum_{1}^{n} f(x_i).$$

- Wenn man annimmt, dass für eine reellwertige Funktion f gilt, dass $f(x)$ nur in kleinen Bereichen von D große Werte annimmt, sonst aber

ungefähr 0 ist, so ist es für die Berechnung des Integrals "verschwendete Zeit" gleichverteilt in den unwichtigen ($f(x) \approx 0$) und den wichtigen Teilen ($f(x) >> 0$) Stichprobenwerte zu ziehen.

- Beispiel: Typische Dichtefunktionen, z.B. Normalverteilung haben die Wahrscheinlichkeitsmasse auf einem kleinen Teil der reellen Achse konzentriert!

- Die Idee ist nun, vorrangig in dem Bereich der Funktion Stichprobenpunkte x_i zu ziehen, für den gilt $f(x) > 0$ und diese Verzerrung nach dem Stichprobenziehen wieder zu korrigieren!

- Die zu erzielende Varianzreduktion ist erheblich.

- Grundlage des Verfahrens ist folgende einfache Beobachtung aus der mathematischen Statistik. Sei p die Dichte einer stetigen ZV X, sowie

eine reellwertige Funktion f gegeben. Sei weiterhin q eine weitere Dichte, aus der man geeignet (einfacher) eine i.i.d. Stichprobe $x_1, \ldots, x_n$ ziehen kann, mit $q(x) > 0$, immer wenn $f(x)p(x) \neq 0$.
Dann gilt mit der Gewichtsfunktion $w(x) := \frac{p(x)}{q(x)}$

$$E_p(f(X)) = \int f(x)p(x)dx = \int w(x)f(x)q(x)dx = E_q(w(X)f(X)).$$

- Damit ist, wenn man unabhängige Zufallszahlen $x_1, \ldots, x_n$ gemäß der Verteilung Q erzeugt hat,

$$f(\bar{X}) = \frac{1}{n}\sum_{1}^{n} w(x_i)f(x_i)$$

ein erwartungstreuer Schätzer für $\int f(x)p(x)$dx.

- Im naiven Ansatz ist $p(x)$ die Dichte der Gleichverteilung, also eine Konstante!

- Ob eine Varianzreduktion erreicht werden kann, hängt davon ab, wie gut man q wählen kann. Oft ist dazu eine Vorabstudie über den Verlauf der Funktion f nötig.

- Könnte man q genau proportional zu f wählen, verschwände die Varianz.

- Dies ist (natürlich) unmöglich, jedoch muss das Ziel sein, eine Dichte zu finden, welche qualitativ möglichst ähnlich zu f verläuft.

- Achtung: Es ist für das *Importance Sampling* wichtig, die Ränder der Verteilung gut zu erwischen!

- Das Verfahren ist nicht auf eindimensionale Dichten beschränkt!

Beispiel (aus Eric Anderson, Lecture Notes for Statistical Genetics)

- Sei $\phi(x)$ die Dichte der Standardnormalverteilung. Angenommen, man möchte

$$\int_{-50}^{50} \phi(x)dx \approx 1$$

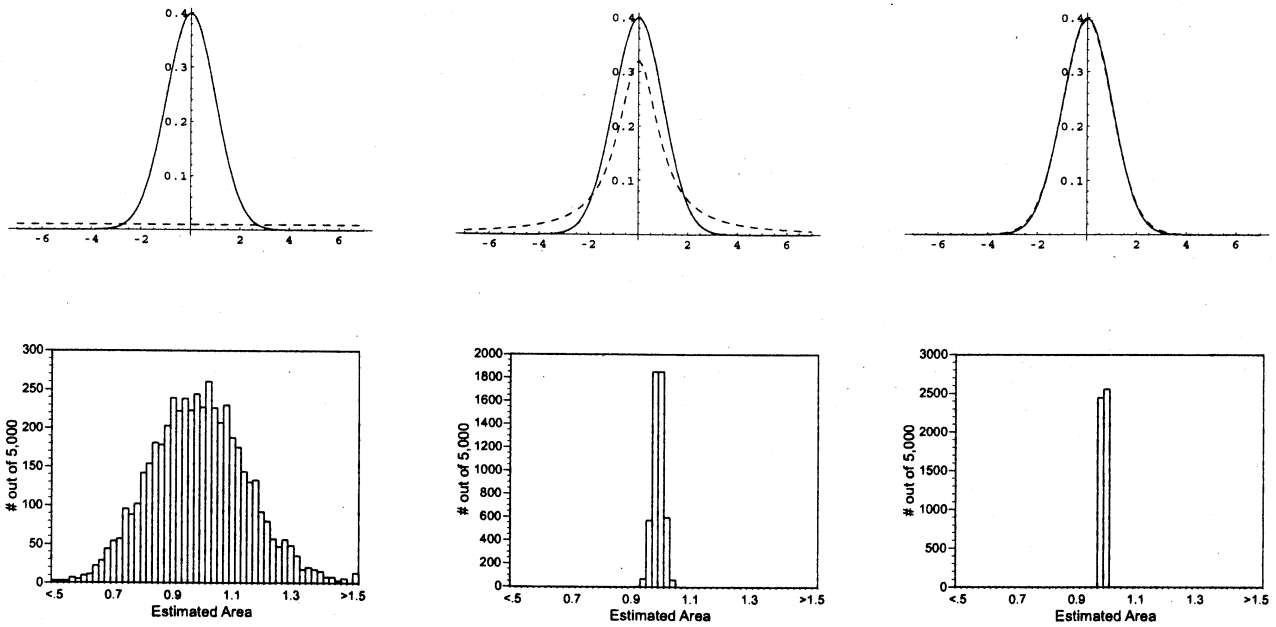
mittels Monte-Carlo-Simulationen berechnen.

- Es sollen drei Ansätze für die approximierenden Dichten verglichen werden:

- Die Gleichverteilung auf [-50, 50], der bekannte Standardansatz, sowie die Dichten der t-Verteilungen t_1 und t_{30}, jeweils auf [-50, 50] gestutzt.

- Es ergeben sich folgende Histogramme für die Verteilungen der Integralschätzer:

Importance-Sampling Schätzer für verschiedende Gewichtsfunktionen



(a) Uniform (b) t_1 Dist. (c) t_30 Dist.

Figure 1: Three different importance sampling functions (dotted lines) used to integrate the standard normal density (solid line) from -50 to 50. Top panels are the density curves and bottom panels are histograms of 5,000 Monte Carlo estimates of the area (which is exactly 1) using $n = 1,000$.

Fazit Monte-Carlo Methoden

- Gleichverteilte Zufallszahlen lassen sich heute mit "Bordmitteln" von R leicht erzeugen. Diese Zahlen erfüllen alle Gütekriterien, die an Pseudo-Zufallszahlen gemeinhin gestellt werden.

- Aus diesen gleichverteilten Zahlen lassen sich beliebige Verteilungen erzeugen. Einige effektive Methoden wurden vorgestellt.

- Viele Aufgaben, die zu schwierig oder zu hochdimensional sind, um sie mit klassischen Verfahren der numerischen Mathematik zu lösen, sind mittels stochastischer Integration lösbar.

- Wichtig ist zu erkennen, dass nur erwartungstreue *Schätzer* für die gesuchten Größen generiert werden.

- Die Varianz der Schätzer (und natürlich der Stichprobenumfang) bestimmt die Ungenauigkeit der Lösungen.

- Schließlich wurden einige Standardverfahren gezeigt, um die Varianz der Schätzer zu reduzieren.