

# Lösungen zu den Aufgaben

## 1. Aufgabe

Erstellen Sie die Posteriori-Verteilung für den Globusversuch. Nutzen Sie dafür diese Syntax:

```
p_grid <- seq( from=0 , to=1 , length.out=1000 ) # Gitterwerte
prior <- rep( 1 , 1000 ) # Priori-Gewichte
likelihood <- dbinom( 6 , size=9 , prob=p_grid )
unstandardisierte_posterior <- likelihood * prior
posterior <- unstandardisierte_posterior / sum(unstandardisierte_posterior)

# um die Zufallszahlen festzulegen, damit alle die gleichen Zufallswerte bekommen:
set.seed(100)

# Stichproben ziehen aus der Posteriori-Verteilung
samples <-
  tibble(
    p = sample( p_grid , prob=posterior, size=1e4, replace=TRUE) %>%
    mutate(
      id = row_number()
    )
  )
```

- Wie viel Wahrscheinlichkeitsmasse liegt unter  $p = 0.2$ ?
- Wie viel Wahrscheinlichkeitsmasse liegt über  $p = 0.8$ ?
- Welcher Anteil der Posteriori-Verteilung liegt zwischen  $p = 0.2$  und  $p = 0.8$ ?
- Unter welchem Wasseranteil  $p$  liegen 10% der Posteriori-Verteilung?
- Über welchem Wasseranteil  $p$  liegen 10% der Posteriori-Verteilung?
- Welches *schmälstes* Intervall von  $p$  enthält 66% der Posteriori-Wahrscheinlichkeit?
- Welcher Wertebereich (synonym: Welches Intervall) von  $p$  enthält 66% der Posteriori-Wahrscheinlichkeit (hier wird *Posteriori-Wahrscheinlichkeit* synonym gebraucht zu *Posteriori-Verteilung*)? Wie nennt man diese Arten von Intervall?

## Lösung

Es finden sich auch Lösungsvorschläge online, z.B. [hier](#)

- Wie viel Wahrscheinlichkeitsmasse liegt unter  $p = 0.2$ ?

```
samples %>%
  filter(p < 0.2) %>%
  summarise(anzahl_zeilen = n(),
            anteil = n() / nrow(samples))
```

```
anzahl_zeilen anteil
4             0
```

Etwas prägnanter könnte der R-Code auch so aussehen:

```
set.seed(100)
stipros <- sample( p_grid , prob=posterior, size=1e4, replace=TRUE)
mean(stipros < 0.2)

## [1] 4e-04
```

- Wie viel Wahrscheinlichkeitsmasse liegt über  $p = 0.8$ ?

```

samples %>%
  filter(p > 0.8) %>%
  summarise(anzahl_zeilen = n(),
            anteil = n() / nrow(samples))

```

```

anzahl_zeilen anteil
      1116  0.11

```

Oder so, prägnanter:

```
mean(stipros>0.8)
```

```
## [1] 0.11
```

c. Welcher Anteil der Posteriori-Verteilung liegt zwischen  $p = 0.2$  und  $p = 0.8$ ?

```

samples %>%
  filter(p > 0.2 & p < 0.8) %>%
  summarise(anzahl_zeilen = n(),
            anteil = n() / nrow(samples))

```

```

anzahl_zeilen anteil
      8880  0.89

```

Oder wieder kürzer geschrieben:

```
mean(stipros > 0.2 & stipros < 0.8)
```

```
## [1] 0.89
```

d. Unter welchem Wasseranteil  $p$  liegen 20% der Posteriori-Verteilung?

Eine Möglichkeit: Wir sortieren  $p$  der Größe nach (aufsteigend), filtern dann so, dass wir nur die ersten 20% der Zeilen behalten und schauen dann, was der größte Wert ist.

```

samples %>%
  arrange(p) %>%
  slice_head(prop = 0.2) %>%
  summarise(quantil_20 = max(p))

```

```

quantil_20
      0.52

```

Andererseits: Das, was wir gerade gemacht haben, nennt man auch ein *Quantil* berechnen, s. auch [hier](#). Dafür gibt's fertige Funktionen in R, wie `quantile()`:

```

samples %>%
  summarise(q_20 = quantile(p, 0.1))

```

```

q_20
      0.45

```

Prägnanter:

```
quantile(stipros, 0.1)
```

```
## 10%
```

```
## 0.45
```

e. Über welchem Wasseranteil  $p$  liegen 10% der Posteriori-Verteilung?

```
quantile(stipros, 0.9)
```

```
## 90%
```

```
## 0.81
```

f. Welches *schmälstes* Intervall von  $p$  enthält 66% der Posteriori-Wahrscheinlichkeit?

```

library(rethinking)
HPDI(stipros, prob = 0.66)

```

```
## |0.66 0.66|
## 0.51 0.77
```

Beachten Sie, dass `HPDI` als Eingabe einen Vektor (bzw. eine Spalte wie etwa `samples$p`) verlangt. Sie könnten die Syntax auch so schreiben, im Tidyverse-Stil:

```
samples %>%
  summarise(hdpi_66 = HPDI(p, prob = 0.66))
```

**hdpi\_66**

0.51

0.77

Alternativ könnte man auch die Funktion `hdi()` aus dem Paket `tidybayes` verwenden:

```
library(tidybayes)
hdi(samples$p, .width = .66)
```

0.51 0.77

g. Welcher Wertebereich von  $p$  enthält 66% der Posteriori-Wahrscheinlichkeit (hier wird *Posteriori-Wahrscheinlichkeit* synonym gebraucht zu *Posteriori-Verteilung*)?

```
library(rethinking)
PI(stipros, prob = 0.66)
```

```
## 17% 83%
## 0.50 0.77
```

Ein "mittleres" 2/3-Intervall lässt 1/3 der Wahrscheinlichkeitsmasse außen vor, und zwar gleichmäßig in zwei Hälften links und rechts, also jeweils 1/6 (17%). So ein Intervall heißt *Perzentilintervall*. Daher:

```
samples %>%
  summarise(PI_66 = quantile(p, prob = c(0.17, .84)))
```

**PI\_66**

0.50

0.77

## 2. Aufgabe

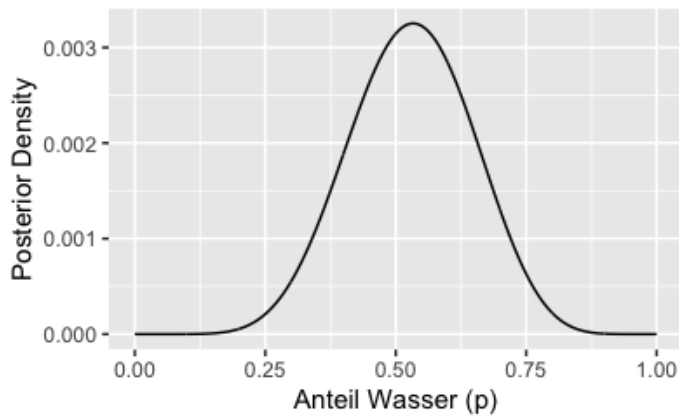
Nehmen wir an, wir haben 8 (Wasser-)“Treffer” ( $W = 8$ ) bei 15 Würfeln ( $N = 15$ ) erhalten (wieder im Globusversuch). Gehen Sie wieder von einer “flachen”, also gleichverteilten, Priori-Verteilung aus.

Berechnen Sie die Posteriori-Verteilung und visualisieren Sie sie. Nutzen Sie die Gittermethode.

## Lösung

```
p_grid <- seq(from = 0, to = 1, length.out = 1000)
prior <- rep(1, 1000)
likelihood <- dbinom(8, size = 15, prob = p_grid)
posterior <- likelihood * prior
posterior <- posterior / sum(posterior)

tibble(p = p_grid, posterior = posterior) %>%
  ggplot(aes(x = p, y = posterior)) +
  # geom_point() +
  geom_line() +
  labs(x = "Anteil Wasser (p)", y = "Posterior Density")
```



[Quelle](#)

### 3. Aufgabe

Nehmen wir an, wir haben 8 (Wasser-)“Treffer” ( $W = 8$ ) bei 15 Würfeln ( $N = 15$ ) erhalten (wieder im Globusversuch).

- Führen Sie einen Posteriori-Prädiktiv-Check durch: Erstellen Sie also eine Posteriori-Prädiktiv-Verteilung (PPV). Mit anderen Worten: Erstellen Sie die Stichprobenverteilung, gemittelt über die Posteriori-Wahrscheinlichkeiten des Wasseranteils  $p$ !
- Visualisieren Sie die PPV!
- Was ist die Wahrscheinlichkeit laut PPV 8 von 15 Treffer zu erzielen (also 8 Wasser in 15 Würfeln)?

### Lösung

a.

Erstellen wir zuerst wieder die Posteriori-Verteilung für den Globusversuch.

```
p_grid <- seq( from=0 , to=1 , length.out=1000 ) # Gitterwerte
prior <- rep(1, 1000 ) # Priori-Gewichte
likelihood <- dbinom(8 , size= 15, prob=p_grid )
unstandardisierte_posterior <- likelihood * prior
posterior <- unstandardisierte_posterior / sum(unstandardisierte_posterior)
```

Dann ziehen wir unsere Stichproben daraus:

```
# um die Zufallszahlen festzulegen, damit alle die gleichen Zufallswerte bekommen:
set.seed(100)

# Stichproben ziehen aus der Posteriori-Verteilung
samples <-
  tibble(
    p = sample(p_grid , prob=posterior, size=1e4, replace=TRUE)) %>%
    mutate(
      id = row_number())

PPV <-
  samples %>%
  mutate( anzahl_wasser = rbinom(1e4, size = 15, prob = p))
```

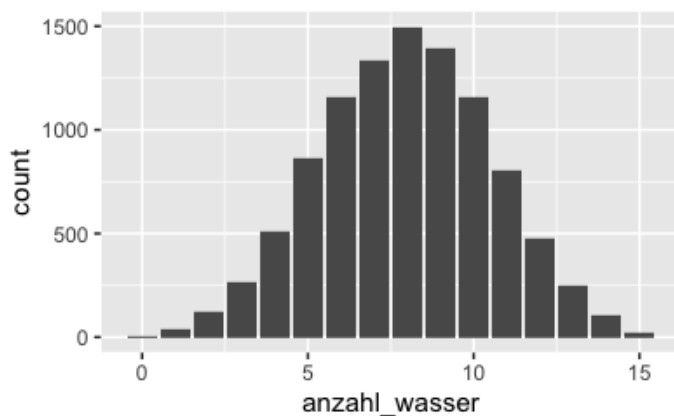
Durch `prob = p` gewichten wir die Wahrscheinlichkeit an den Werten der Posteriori-Verteilung.

So sehen die ersten paar Zeilen von `PPV` aus:

p	id	anzahl_wasser
0.47	1	5
0.36	2	2
0.48	3	6
0.47	4	7
0.50	5	7

b.

```
PPV %>%  
  ggplot() +  
  aes(x = anzahl_wasser) +  
  geom_bar()
```



c.

```
PPV %>%  
  filter(anzahl_wasser == 8) %>%  
  summarise(wskt_wasser = n() / nrow(PPV))
```

```
wskt_wasser  
0.15
```

Prägnanterer R-Code:

```
w <- rbinom(1e4, size = 15, prob = samples$p)  
mean(w == 8)  
## [1] 0.14
```

[Quelle](#)

#### 4. Aufgabe

Nehmen wir an, wir haben 8 (Wasser-)“Treffer” ( $W = 8$ ) bei 15 Würfeln ( $N = 15$ ) erhalten (wieder im Globusversuch).

Berechnen Sie auf Basis der aktuellen Posteriori-Verteilung die Wahrscheinlichkeit für 6 Wasser bei 9 Würfeln ( $W = 6, N = 9$ ).

**Lösung**

Erstellen wir zuerst wieder die Posteriori-Verteilung für den Globusversuch.

```
p_grid <- seq( from=0 , to=1 , length.out=1000 ) # Gitterwerte
prior <- rep( 1 , 1000 ) # Priori-Gewichte
likelihood <- dbinom( 6 , size=9 , prob=p_grid )
unstandardisierte_posterior <- likelihood * prior
posterior <- unstandardisierte_posterior / sum(unstandardisierte_posterior)
```

Dann ziehen wir unsere Stichproben daraus:

```
# um die Zufallszahlen festzulegen, damit alle die gleichen Zufallswerte bekommen:
set.seed(100)

# Stichproben ziehen aus der Posteriori-Verteilung
samples <-
  tibble(
    p = sample( p_grid , prob=posterior, size=1e4, replace=TRUE) %>%
    mutate(
      id = row_number())

PPV <-
  samples %>%
  mutate(anzahl_wasser2 = rbinom(1e4, size = 9, prob = p))

PPV %>%
  filter(anzahl_wasser2 == 6) %>%
  summarise(wie_offt_6_wasser = n(),
            anteil_6_wasser = wie_offt_6_wasser / nrow(samples))
```

<b>wie_offt_6_wasser</b>	<b>anteil_6_wasser</b>
1941	0.19

[Quelle](#)

## 5. Aufgabe

Nehmen wir an, wir haben 8 (Wasser-)“Treffer” ( $W = 8$ ) bei 15 Würfeln ( $N = 15$ ) erhalten (wieder im Globusversuch).

Nehmen Sie dieses Mal *keine* gleichverteilte Priori-Verteilung an. Stattdessen verwenden Sie einen Priori-Wert von Null solange  $p < 0.5$  und einen konstanten Wert für  $p \geq 0.5$ . Diese Priori-Verteilung kodiert die Information, dass mindestens die Hälfte der Erdoberfläche mit Sicherheit aus Wasser besteht.

Für alle folgenden Berechnungen, vergleichen Sie Ihre Ergebnisse zu der analogen Analyse mit einem konstanten (gleichverteilten) Priori-Wert!

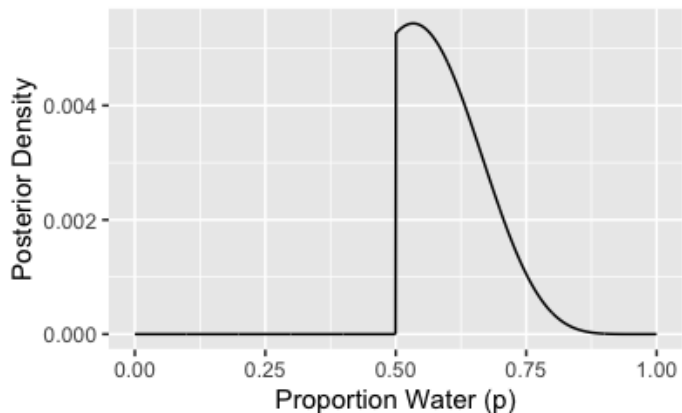
- Berechnen Sie die Posteriori-Verteilung und visualisieren Sie sie. Nutzen Sie die Gittermethode.
- Ziehen Sie  $10^4$  Stichproben aus der Posteriori-Verteilung, die Sie mit der Gittermethode erhalten haben. Berechnen Sie auf dieser Grundlage das 90%-HDPI.
- Berechnen Sie die PPV für dieses Modell. Was ist die Wahrscheinlichkeit 8 von 15 Treffer zu erzielen laut dieser PPV?
- Auf Basis der aktuellen Posteriori-Wahrscheinlichkeit: Was ist die Wahrscheinlichkeit für 6 Wasser bei 9 Würfeln?

## Lösung

- Berechnen Sie die Posteriori-Verteilung und visualisieren Sie sie. Nutzen Sie die Gittermethode.

```
p_grid <- seq(from = 0, to = 1, length.out = 1000)
prior <- case_when(
  p_grid < 0.5 ~ 0,
  p_grid >= 0.5 ~ 1)
likelihood <- dbinom(8, size = 15, prob = p_grid)
unstand_posterior <- likelihood * prior
posterior <- unstand_posterior / sum(unstand_posterior)
```

```
tibble(p = p_grid,
       posterior = posterior) %>%
  ggplot(aes(x = p, y = posterior)) +
  # geom_point() +
  geom_line() +
  labs(x = "Proportion Water (p)", y = "Posterior Density")
```



- b. Ziehen Sie  $10^4$  Stichproben aus der Posteriori-Verteilung, die Sie mit der Gittermethode erhalten haben. Berechnen Sie auf dieser Grundlage das 90%-HDPI.

```
library(rethinking)
# Stichproben (samples) aus der Posteriori-Verteilung:
samples <- sample(p_grid, prob = posterior, size = 1e4, replace = TRUE)
HPDI(samples, prob = 0.9)

## |0.9 0.9|
## 0.50 0.71
```

- c. Berechnen Sie die PPV für dieses Modell. Was ist die Wahrscheinlichkeit 8 von 15 Treffer zu erzielen laut dieser PPV?

```
PPV <-
  tibble(w = rbinom(1e4, size = 15, prob = samples)) # w wie Wasser

PPV %>%
  filter(w == 8) %>%
  summarise(anteil_8_wasser = n()/nrow(PPV))
```

**anteil\_8\_wasser**

0.15

- d. Auf Basis der aktuellen Posteriori-Wahrscheinlichkeit: Was ist die Wahrscheinlichkeit für 6 Wasser bei 9 Würfeln?

```
PPV <-
  PPV %>%
  mutate(w2 = rbinom(1e4, size = 9, prob = samples))

PPV %>%
  filter(w2 == 6) %>%
  summarise(anteil_6_wasser = n()/nrow(PPV))
```

**anteil\_6\_wasser**

0.22

[Quelle](#)

## 6. Aufgabe

Nach einem langen Unitag machen Sie sich auf den Weg nach Hause; ihr Weg führt Sie durch eine dunkle Ecke. Just dort regt sich auf einmal eine Gestalt in den Schatten. Die Person spricht Sie an: „Na, Lust auf ein Spielchen?“. Sie willigen sofort ein. Die Person stellt sich als ein Statistiker vor, dessen Namen nichts zur Sache tue; das Gesicht kommt Ihnen vage bekannt vor. „Pass auf“, erklärt der Statistiker, „wir werfen eine Münze, ich setze auf Zahl“. Dass er auf Zahl setzt, überrascht Sie nicht. „Wenn ich gewinne“, fährt der Statistiker fort, „bekomme ich 10 Euro von Dir, wenn Du gewinnst, bekommst Du 11 Euro von mir. Gutes Spiel, oder?“. Sie einigen sich auf 10 Durchgänge, in denen der Statistiker jedes Mal eine Münze wirft, fängt und dann die oben liegende Seite prüft. Erster Wurf: Zahl! Der Statistiker gewinnt. Pech für Sie. Zweiter Wurf: Zahl! Schon wieder 10 Euro für den Statistiker. Hm. Dritter Wurf: . . . Zahl! Schon wieder. Aber kann ja passieren, bei einer fairen Münze, oder? Vierter Wurf: Zahl! Langsam regen sich Zweifel bei Ihnen. Kann das noch mit rechten Dingen zugehen? Ist die Münze fair? Insgesamt gewinnt der zwielichte Statistiker 8 von 10 Durchgängen.

Unter leisem Gelächter des Statistikers (und mit leeren Taschen) machen Sie sich von dannen. Hat er falsch gespielt? Wie plausibel ist es, bei 10 Würfeln 8 Treffer zu erhalten, wenn die Münze fair ist? Ist das ein häufiges, ein typisches Ereignis oder ein seltenes, untypisches Ereignis bei einer fairen Münze? Wenn es ein einigermaßen häufiges Ereignis sein sollte, dann spricht das für die Fairness der Münze. Zumindest spricht ein Ereignis, welches von einer Hypothese als häufig vorausgesagt wird und schließlich eintritt, nicht gegen eine Hypothese. Zuhause angekommen, denken Sie sich, jetzt müssen Sie erstmal in Ruhe die Posteriori-Verteilung und die PPV ausrechnen!

- Berechnen Sie die Posteriori-Verteilung mit der Gittermethode! Gehen Sie von einer gleichverteilten Priori-Wahrscheinlichkeit aus. Visualisieren Sie sie. Alle folgenden Teil-Fragen bauen auf der Post-Verteilung auf.
- Wie groß ist die Wahrscheinlichkeit, auf Basis der Post-Verteilung, dass die Münze zugunsten des Dozenten gezinkt ist?
- Geben Sie das 50%-PI und 50%-HDPI zum Parameterwert ( $p$  der Münze) an!
- Mit welcher Wahrscheinlichkeit liegt die Trefferchance der Münze zwischen  $p = .45$  und  $p = .55$ , ist also nicht „nennenswert“ gezinkt?
- Was ist der wahrscheinlichste Parameterwert (Trefferchance der Münze)?
- Geben Sie das 90%-PI und 90%-HDPI zu Parameterwert ( $p$  der Münze) an!
- Berechnen Sie die PPV! Visualisieren Sie sie. Interpretieren Sie die PPV.
- Diskutieren Sie die Annahme einer Gleichverteilung des Priori-Wertes von  $p$ !

## Lösung

- Berechnen Sie die Posteriori-Verteilung mit der Gittermethode! Visualisieren Sie sie. Alle folgenden Teil-Fragen bauen auf der Post-Verteilung auf.

```
p_grid <- seq( from=0 , to=1 , length.out=1000 ) # Gitterwerte
prior <- rep( 1 , 1000 ) # Priori-Gewichte
likelihood <- dbinom(8, size = 10, prob=p_grid)
unstandardisierte_posterior <- likelihood * prior
posterior <- unstandardisierte_posterior / sum(unstandardisierte_posterior)

# Stichproben ziehen aus der Posteriori-Verteilung:
samples <-
  tibble(
    gewinnchance_muenze = sample(p_grid , prob=posterior, size=1e4, replace=TRUE) %>%
    mutate(
      id = row_number()
    )
  )
```

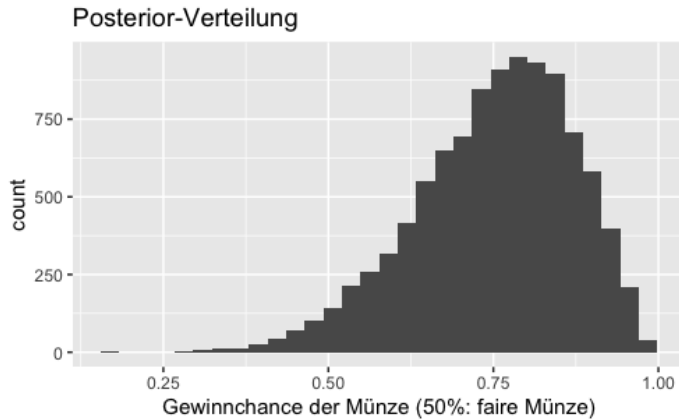
Visualisierung:



```

samples %>%
  ggplot() +
  aes(x = gewinnchance_muenze) +
  geom_histogram() +
  labs(title = "Posterior-Verteilung",
        x = "Gewinnchance der Münze (50%: faire Münze)")

```



b. Wie groß ist die Wahrscheinlichkeit, auf Basis der Post-Verteilung, dass die Münze zugunsten des Dozenten gezinkt ist?

```

samples %>%
  filter(gewinnchance_muenze > 0.5) %>%
  summarise(wskt_gezinkt = n()/nrow(samples))

```

**wskt\_gezinkt**  
0.97

`n()` gibt die Anzahl der Zeilen im aktuellen Tabelle zurück, also nach dem Filtern. `nrow()` gibt die Anzahl der Zeilen in der ursprünglichen Tabelle zurück.

c. Geben Sie das 50%-PI und 50%-HPDI zum Parameterwert ( $p$  der Münze) an!

```

library(rethinking)
samples %>%
  summarise(PI = PI(gewinnchance_muenze, prob = .5),
            HPDI = HPDI(gewinnchance_muenze, prob = .5))

```

**PI HPDI**  
0.68 0.71  
0.84 0.87

d. Mit welcher Wahrscheinlichkeit liegt die Trefferchance der Münze zwischen  $p = .45$  und  $p = .55$ , ist also nicht "nennenswert" gezinkt?

```

samples %>%
  filter(gewinnchance_muenze >= 0.45 & gewinnchance_muenze <= .55) %>%
  summarise(wskt_nichtgezinkt = n()/nrow(samples))

```

**wskt\_nichtgezinkt**  
0.05

e. Was ist der wahrscheinlichste Parameterwert (Trefferchance der Münze)?

```

library(tidybayes)
samples %>%
  summarise(modus = Mode(gewinnchance_muenze))

```

**modus**  
0.78

f. Geben Sie das 90%-PI und 90%-HPDI zu Parameterwert ( $p$  der Münze) an!

```

#library(rethinking)
samples %>%
  summarise(PI = PI(gewinnchance_muenze, prob = .9),
            HPDI = HPDI(gewinnchance_muenze, prob = .9))

```

## PI HPDI

0.53 0.57

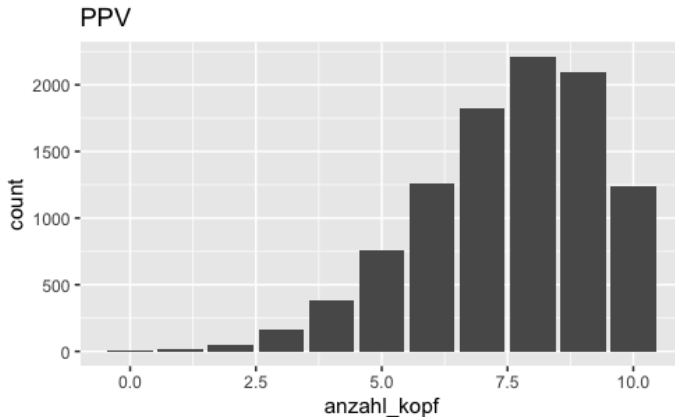
0.92 0.94

g. Berechnen Sie die PPV! Visualisieren Sie sie. Interpretieren Sie die PPV.

```
PPV <-  
  samples %>%  
  mutate(anzahl_kopf = rbinom(n = 1e4, size = 10, prob = gewinnchance_muenze))
```

Visualisierung:

```
PPV %>%  
  ggplot() +  
  aes(x = anzahl_kopf) +  
  labs(title = "PPV") +  
  geom_bar() # geom_bar() ginge auch, sieht aber bei wenig Balken nicht so gut aus.
```



Laut der PPV sind 8 von 10 Treffern der Wert, der mit der höchsten Wahrscheinlichkeit zu beobachten sein wird. Allerdings sind 7 oder 9 Treffer fast genauso wahrscheinlich. Etwas genauer:

```
PPV %>%  
  filter(between(anzahl_kopf, 7, 9)) %>% # "filter mir einen Wert ZWISCHEN 7 und 9 ..."  
  summarise(anteil_7bis9 = n()/nrow(PPV))
```

## anteil\_7bis9

0.61

Mit dieser Wahrscheinlichkeit ist ein Wert zwischen 7 und 9 zu beobachten, wenn man den Versuch wiederholt, laut dem Modell.

```
PPV %>%  
  summarise(pi = PI(anzahl_kopf, prob = .9))
```

## pi

4

10

Unser Modell sieht einen "Passungsbereich" (ein Perzentilintervall) von 4 bis 10 Treffern als mit 90% Wahrscheinlichkeit passend an.

h. Diskutieren Sie die Annahme einer Gleichverteilung des Priori-Wertes von  $p$ !

Zwar hat eine Gleichverteilung der Priori-Werte den Vorteil, dass sie "objektiv" ist in dem Sinne, dass kein Wert "bevorzugt" wird; alle gelten als gleich wahrscheinlich. Aber das ist hochgradig unplausibel: So ist z.B. der Wert  $p = 1$  logisch unmöglich, da wir nicht nur Treffer beobachtet haben. Ein Wert von z.B.  $p = 0.999$  erscheint uns ebenfalls sehr unwahrscheinlich. Nützlicher erscheint daher vielleicht doch eine Priori-Verteilung, die extreme Werte von  $p$  als unwahrscheinlich bemisst.

## 7. Aufgabe

Sie sind kürzlich in ein Startup-Unternehmen eingestiegen. Das Unternehmen versucht, einen Online-Weinhandel aufzubauen. Kern des Unternehmens ist eine künstliche Intelligenz, die versucht, den Kundis den best möglich passenden Wein anzudreh... zu verkaufen.

Sie haben sich bei Ihrem Bewerbungsgespräch persönlich von der Qualität der Produkte eingehend überzeugt und sind daher hoch motiviert, sich zum Wohle des Unternehmens einzusetzen.

Kürzlich hat eine Beratungsfirma, die Ihre Kunden im Rahmen einer qualitativen Studie untersucht hat, herausgefunden, dass doch ein beachtlicher Teil von einem Menschen, nicht von einem Roboter (bzw. der KI) beim Wein aussuchen beraten werden möchte. Diesen Anteil von Kunden (die nicht von der KI beraten werden möchten) möchten Sie jetzt genauer bestimmen.

Dazu haben Sie  $N = 42$  Kundis befragt. Gut die Hälfte ( $n = 23$ ) hat sich zugunsten der KI ausgesprochen; der Rest der Kundis möchte lieber von einem Menschen beraten werden.

Gehen Sie im Folgenden davon aus, dass die Studie bzw. die erhaltenen Daten von guter Qualität ist (man also keine Probleme wie mangelnde Repräsentativität erwarten muss).

Verwenden Sie die Gittermethode und gleichverteilte Priori-Werte.

- Wie groß ist die Wahrscheinlichkeit, dass die KI-freundlichen Kundis bei Ihnen überwiegen?
- Wie groß ist die Wahrscheinlichkeit (laut Modell), dass künftig eine Mehrheit an KI-freundlichen Kundis zu *beobachten* sein wird?
- Wenn Sie nur eine Zahl angeben dürften: Was ist Ihr Schätzwert zum Anteil der KI-Freunde (in dieser Studie)?

## Lösung

- Wie groß ist die Wahrscheinlichkeit (laut Modell), dass die KI-freundlichen Kundis bei Ihnen überwiegen?

Das ist eine Frage nach dem *inversen Quantil* (synonym: *kumulative Verteilungsfunktion*, cumulative distribution function, cdf).

```
p_grid <- seq(from=0,
             to=1,
             length.out=1000) # Gitterwerte

prior <- rep(1, 1000) # Priori-Gewichte

set.seed(42) # Zufallszahlen festlegen
likelihood <- dbinom(23, size = 42, prob=p_grid)

unstandardisierte_posterior <- likelihood * prior

posterior <- unstandardisierte_posterior / sum(unstandardisierte_posterior)
```

Ziehen wir daraus Stichproben:

```
set.seed(42) # Zufallszahlen festlegen
samples <-
  tibble(
    p = sample(p_grid,
              prob = posterior,
              size=1e4,
              replace=TRUE) %>%
    mutate(id = 1:nrow(samples))

## Error: Problem with `mutate()` column `id`.
## i `id = 1:nrow(samples)`.
## x object 'samples' not found

samples %>%
  filter(p > 0.5) %>%
  summarise(wskt_mehrheit_will_ki = n()/nrow(samples))
```

```
## Error in filter(., p > 0.5): object 'samples' not found
```

```
samples %>%  
  ggplot() +  
  aes(x = p) +  
  geom_histogram() +  
  geom_vline(xintercept = 0.5) +  
  labs(title = "Post-Verteilung")
```

```
## Error in ggplot(.): object 'samples' not found
```

**b. Wie groß ist die Wahrscheinlichkeit (laut Modell), dass künftig eine Mehrheit an KI-freundlichen Kuffis zu beobachten sein wird?**

```
PPV <-  
  samples %>%  
  mutate(Anzahl_will_KI = rbinom(n = 1e4, size = 42, prob = p))
```

```
## Error in mutate(., Anzahl_will_KI = rbinom(n = 10000, size = 42, prob = p)): object 'samples' not found
```

```
PPV %>%  
  ggplot() +  
  aes(x = Anzahl_will_KI) +  
  geom_histogram() +  
  labs(title = "PPV")
```

```
## Error in ggplot(.): object 'PPV' not found
```

**Eine Mehrheit entspricht mind. 22 von 42 Personen.**

```
PPV %>%  
  filter(Anzahl_will_KI >= 22) %>%  
  summarise(prob_mehrheit_will_ki = n()/nrow(PPV))
```

```
## Error in filter(., Anzahl_will_KI >= 22): object 'PPV' not found
```

**c. Wenn Sie nur eine Zahl angeben dürften: Was ist Ihr Schätzwert zum Anteil der KI-Freunde (in dieser Studie)?**

**Man könnte den Mittelwert oder den Median angeben:**

```
library(rstatix)  
get_summary_stats(samples)
```

```
## Error in is_grouped_df(data): object 'samples' not found
```