Clase 5: Pruebas de hipótesis a la Bayes

Dr. Héctor Nájera
Dr. Curtis Huffman

Convenciones respecto a *Ho:*

- H: El cambio anual en el salario mínimo de referencia tiene un efecto negativo sobre el riesgo de vivir en pobreza multidimensional ($\theta < 0$)

- Ho: $\theta = 0$

- De manera que si p-value<.05 rechazamos Ho y concluimos que probablemente $\theta < 0$

- Publicamos como si hubieramos encontrado evidencia de que descubrirmos algo

- No sólo eso, si $\theta < .315$. Hacemos afirmaciones de tamaño de efecto.

Convenciones

- H: El salari … n efecto negativo sobre … r en pobreza multidim …
- Ho: θ …
- De m … e si p-value<.05 rec … oncluimos que … emente $\theta < 0$ …
- Publican … hubieramos encontrado ev … rirmos algo
- No sólo eso, si … efecto.

La prueba de significancia bajo la hipótesis nula (NHTS)

Ho: $\theta = 0$

Función de verosimilitud (Bernoulli).
Genero I bajo que la nula es cierta

N=ENIGH 2022

Z=Modelo de regresión

Evalúo si el valor obtenido de $\hat{\theta}$ es extremo o no.

La prueba de significancia bajo la hipótesis nula (NHTS)

Ho: $\theta = 0$

Función de verosimilitud (Bernoulli). Genero "I" (nube) bajo que la nula es cierta usando N y Z

N=ENIGH 2022

Z=Modelo de regresión

Z=Modelo de regresión 2

Evalúo si el valor obtenido de $\hat{\theta}$ es extremo o no.

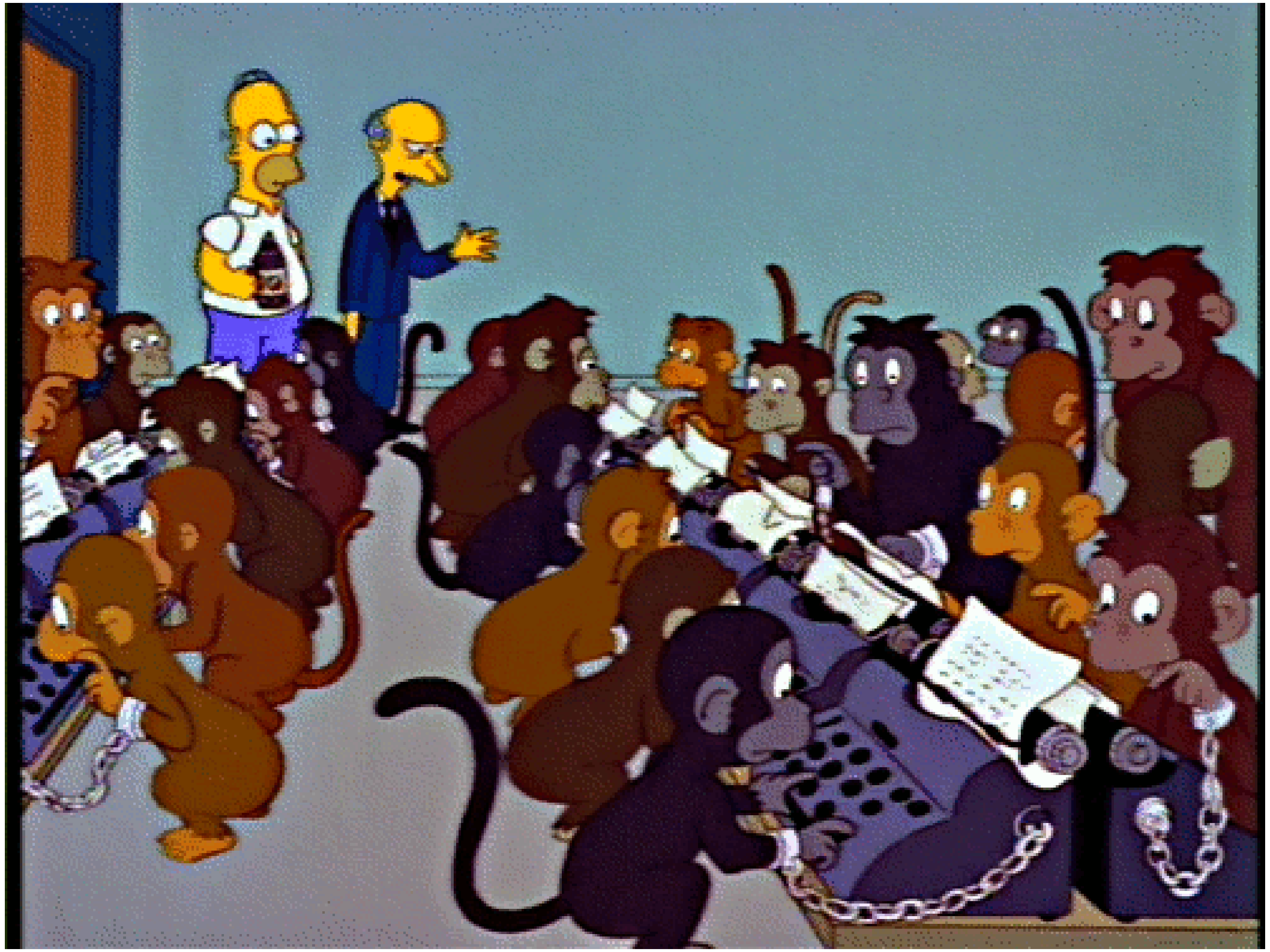
Interpretación correcta bajo NHTS

- H: El cambio anual en el salario mínimo de referencia tiene un efecto negativo sobre el riesgo de vivir en pobreza multidimensional ($\theta < 0$)

$\theta = .345$ (p<.05)

En un mundo en el que el salario mínimo de referencia NO TIENE un efecto sobre el riesgo de vivir en pobreza, hay menos de 5% de probabilidad de observar que no hay diferencias.

Ho: Los datos fueron generados por una máquina generadora de números



La prueba de significancia bajo la hipótesis nula (NHTS)

- La hipótesis nula NO es un modelo científico. Es una técnica de generación numérica aleatoria
- El p-value NO es la probabilidad de que la Ho sea cierta. Es la probabilidad de ver un estadístico distinto condicional en que los datos hayan venido de un generador de números aleatorio
- El p-value depende de lo que hubiéramos hecho bajos otros posibles datos.
- El p-value, en principio, funcionaría bajo condiciones en las que puedo calcular "I" con todos los escenarios posibles (Z) y una N grande (Poder estadístico suficiente)

Paradoja de Meehl

- Si tuviéramos mayor precisión deberíamos rechazar más hipótesis (no nulas): Es más fácil equivocarse

- El efecto es el opuesto. Tenemos más precisión, pero rechazamos con mayor facilidad la hipótesis nula.

- Es como si rechazar la nula fuera muy fácil

Philosophy of Science, 1967, Vol. 34, 103–115. #74

Philosophy of Science

June, 1967

THEORY-TESTING IN PSYCHOLOGY AND PHYSICS: A METHODOLOGICAL PARADOX*

PAUL E. MEEHL[1]

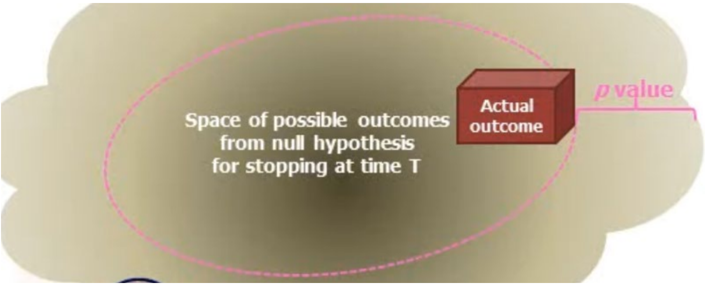
Minnesota Center for Philosophy of Science

Because physical theories typically predict numerical values, an improvement in experimental precision reduces the tolerance range and hence increases corroborability. In most psychological research, improved power of a statistical design leads to a prior probability approaching ½ of finding a significant difference in the theoretically predicted direction. Hence the corroboration yielded by "success" is very weak, and becomes weaker with increased precision. "Statistical significance" plays a logical role in psychology precisely the reverse of its role in physics. This problem is worsened by certain unhealthy tendencies prevalent among psychologists, such as a premium placed on experimental "cuteness" and a free reliance upon *ad hoc* explanations to avoid refuation.

Solución a la paradoja de Meehl

- Buscar aceptar el valor que se predice y no buscar rechazar un valor nulo no predicho

- Esto significa aumentar la precisión de la predicción del valor predicho



AH!

- New statistics: P-values<.001, Tamaños de efecto, intervalos de confianza y meta análisis como medios para evitar los problemas asociados a pruebas de significancia de las hipótesis nula (NHST).

- P-value: Es la probabilidad de que el estadístico (t) su hubiera obtenido de la nula (dada cierta regla respecto al tamaño de los resultados posibles).

American statistical association (ASA)

- The ASA's statement on p-values says, "*Valid scientific conclusions based on p-values and related statistics cannot be drawn without at least knowing how many and which analyses were conducted.*"
- The whole point of the "garden of forking paths" (Gelman and Loken 2014) is that to compute a valid p-value you need to know what analyses would have been done had the data been different
- The concern is not just multiple comparisons, it is multiple potential comparisons

The Problems With P-Values are not Just With P-Values

Andrew GELMAN

The ASA's statement on p-values says, "Valid scientific conclusions based on p-values and related statistics cannot be drawn without at least knowing how many and which analyses were conducted." I agree, but knowledge of how many analyses were conducted etc. is not enough. The whole point of the "garden of forking paths" (Gelman and Loken 2014) is that to compute a valid p-value you need to know what analyses *would have been done* had the data been different. Even if the researchers only did a single analysis of the data at hand, they well could've done other analyses had the data been different. Remember that "analysis" here also includes rules for data coding, data exclusion, etc.

When I was sent an earlier version of the ASA's statement, I suggested changing the sentence to, "Valid p-values cannot be drawn without knowing, not just what was done with the existing data, but what the choices in data coding, exclusion, and analysis would have been, had the data been different. This credible intervals, Bayes factors, cross-validation: you name the method, it can and will be twisted, even if inadvertently, to create the appearance of strong evidence where none exists.

What, then, can and should be done? I agree with the ASA statement's final paragraph, which emphasizes the importance of design, understanding, and context—and I would also add measurement to that list.

What went wrong? How is it that we know that design, data collection, and interpretation of results in context are so important—and yet the practice of statistics is so associated with p-values, a typically misused and misunderstood data summary that is problematic even in the rare cases where it can be mathematically interpreted?

I put much of the blame on statistical education, for two reasons.

Pruebas de hipótesis con Bayes

- H: El cambio anual en el salario mínimo de referencia tiene un efecto negativo sobre el riesgo de vivir en pobreza multidimensional ($\theta < 0$)

- Prior ¿Cuál es el prior aquí? ¿Cómo se ve? ¿Dónde vive?

Pasos en inferencia bayesiana

Hipótesis

Pasar a parámetros mi hipótesis

Modelo generativo
(Verosimilitud+Datos+Priors)

Me interesa la distribución posterior del parámetro de mi hipótesis

Próximamente…

Pruebas de hipótesis con Bayes

- H: El cambio anual en el salario mínimo de referencia tiene un efecto negativo sobre el riesgo de vivir en pobreza multidimensional ($\theta < 0$)

- Prior ¿Cuál es el prior aquí? ¿Cómo se ve? ¿Dónde vive?

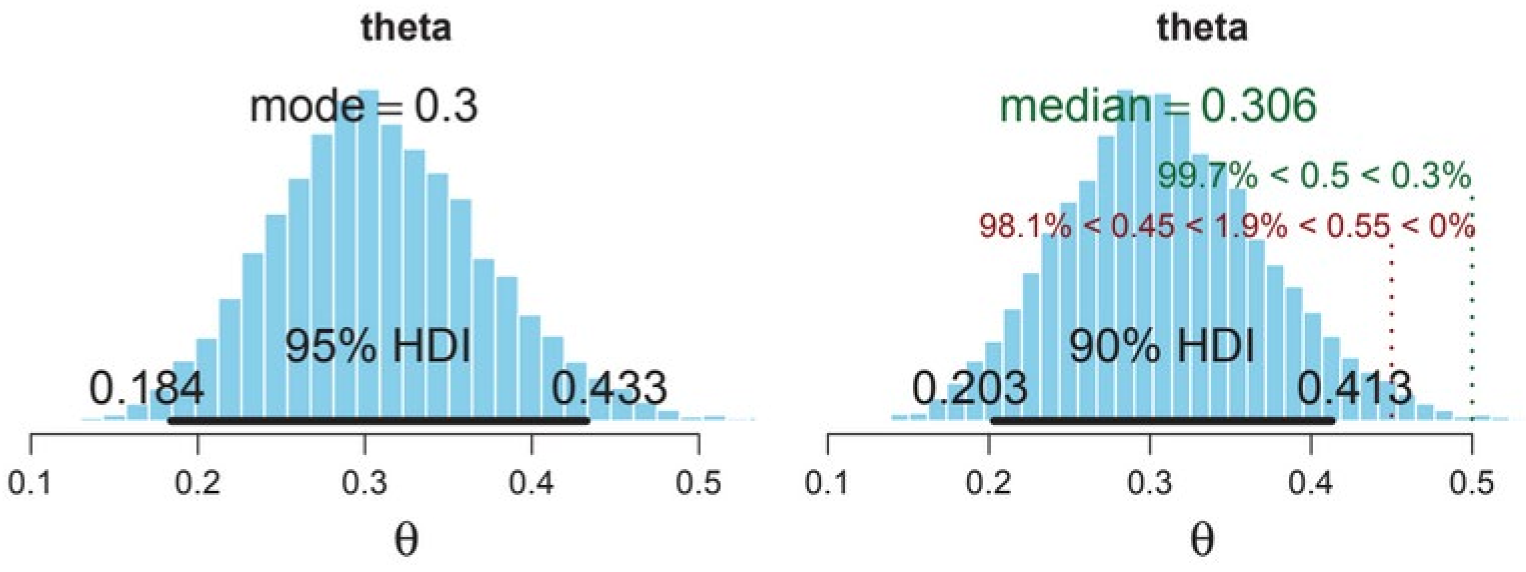
- Encontramos: $\theta = .347$

¿Qué probabilidad hay de que el efecto sea negativo y que ronde por .34?

La precisión en comparaciones es la meta

- HDI: Medición de la precisión del valor posterior del parámetro

- Intervalo de alta densidad (HDI)



Algunas consideraciones extra de la inferencia bayesiana

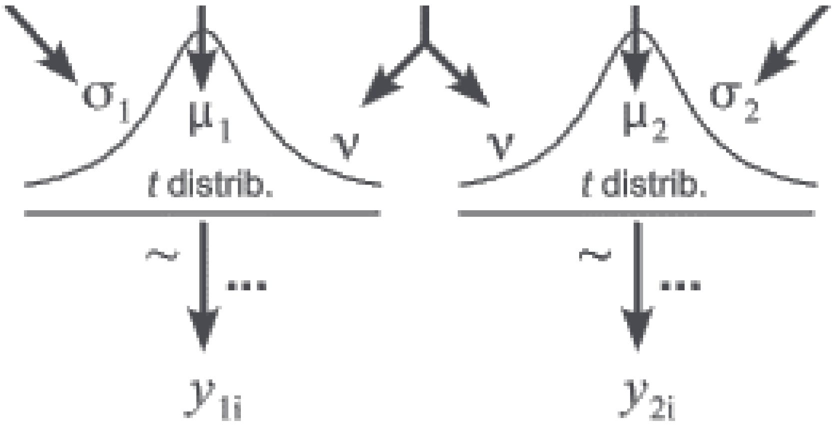
Comparación de parámetros

- Queremos comparar las medias/parámetros entre dos poblaciones
- Quisiéramos saber si hay evidencia de que una población tiene una media más alta que otra
- Tenemos como datos: y1 & y2
- Lo que nos interesa son los parámetros: $\mu_1 > \mu_2$

Bayes: Modelos generativos

Las distribuciones observadas de y1 y y2 provienen de otras distribuciones:

Nos interesa saber:
μ_1 μ_2

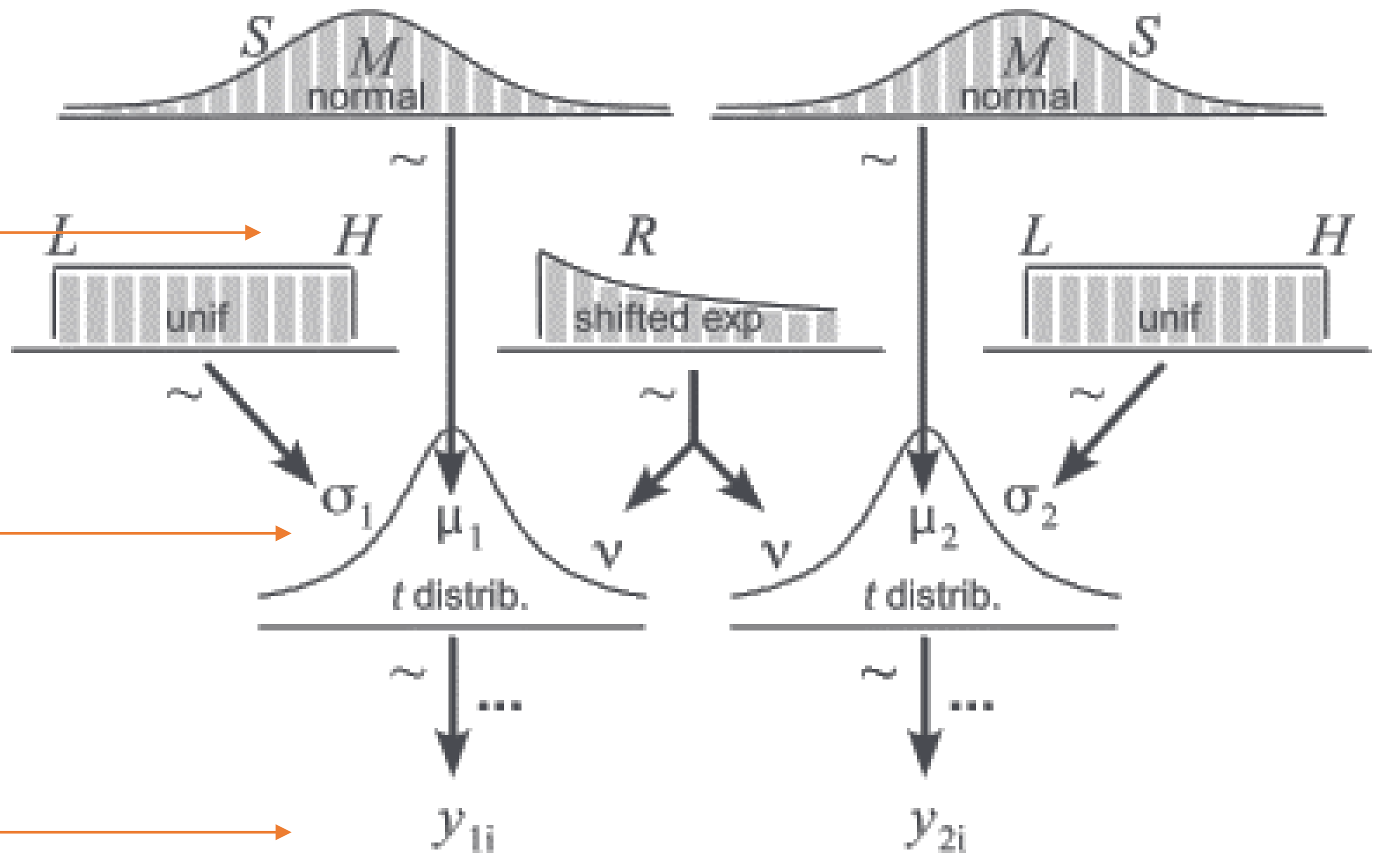


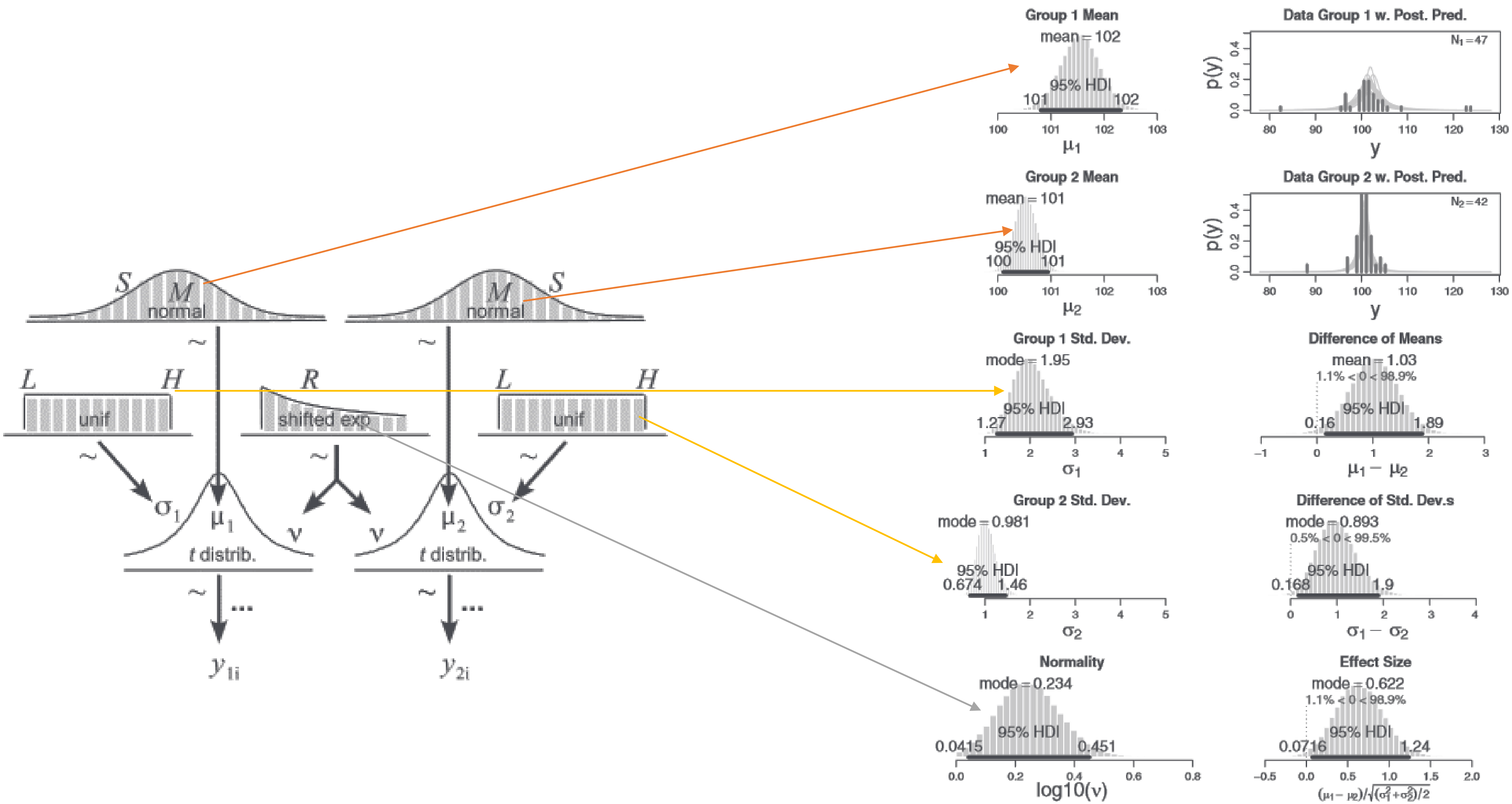
Hipótesis: Diferencias entre dos grupos

Priors: ¿Qué sabemos de cada parámetro antes de estos datos?

Los datos de cada grupo se describen con distribuciones t usando parámetros: mu, nu y sigma

Los datos “fijos”




S
M
normal
S
M
normal
L
H
unif
R
shifted exp
L
H
unif
σ1
μ1
ν
ν
μ2
σ2
t distrib.
t distrib.
...
...
y1i
y2i
Group 1 Mean
mean = 102
95% HDI
101
102
μ1
Data Group 1 w. Post. Pred.
N1 = 47
p(y)
y
Group 2 Mean
mean = 101
95% HDI
100
101
μ2
Data Group 2 w. Post. Pred.
N2 = 42
p(y)
y
Group 1 Std. Dev.
mode = 1.95
95% HDI
1.27
2.93
σ1
Difference of Means
mean = 1.03
1.1% < 0 < 98.9%
95% HDI
0.16
1.89
μ1 − μ2
Group 2 Std. Dev.
mode = 0.981
95% HDI
0.674
1.46
σ2
Difference of Std. Dev.s
mode = 0.893
0.5% < 0 < 99.5%
95% HDI
0.168
1.9
σ1 − σ2
Normality
mode = 0.234
95% HDI
0.0415
0.451
log10(ν)
Effect Size
mode = 0.622
1.1% < 0 < 98.9%
95% HDI
0.0716
1.24
(μ1 − μ2)/√((σ1² + σ2²)/2)

Combinando las posteriores

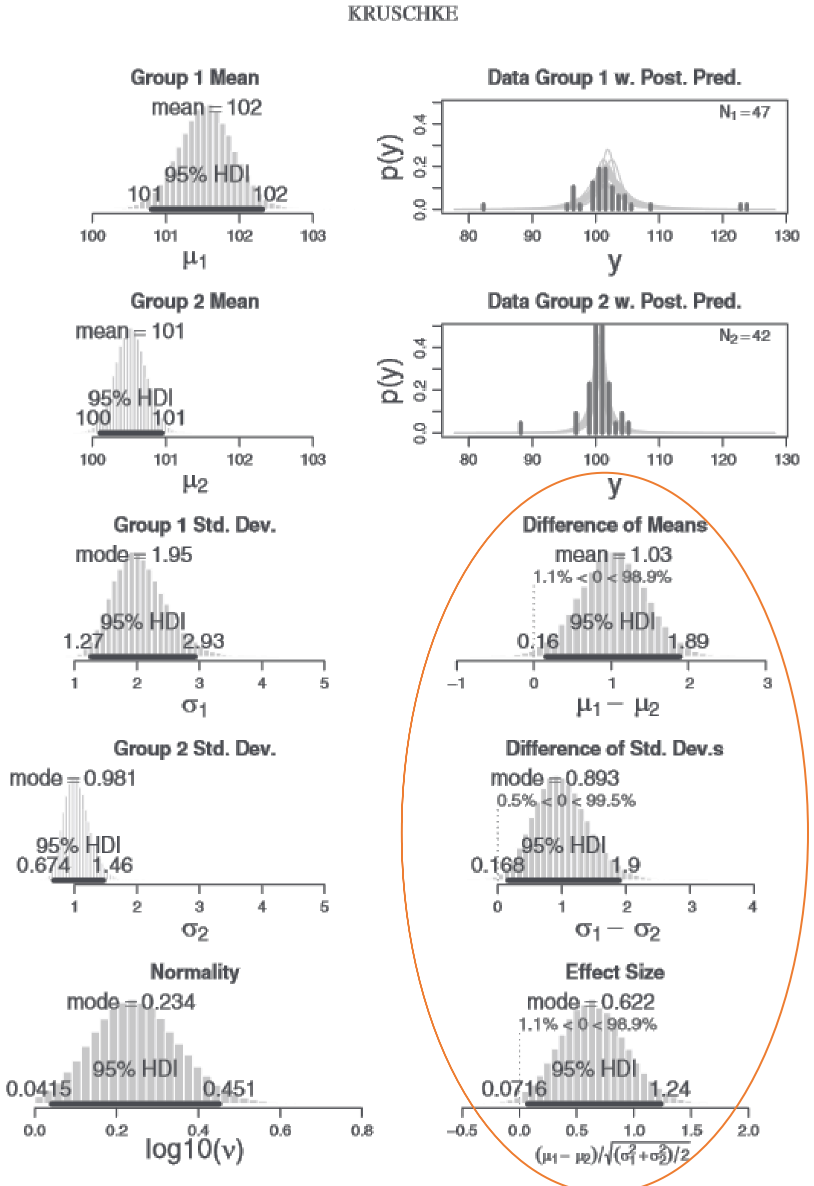
El computo bayesiano ofrece distribuciones de los valores posibles de los parámetros de interés:

Diferencia de medias, de escala (dispersión) y del tamaño del efecto

$$\mu_1 - \mu_2$$

$$\sigma_1 - \sigma_2$$

$$\left(\mu_1 - \mu_2 \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}\right)$$



Manipulaciones con las posteriores

La distribución posterior se puede sumar, restar, multiplicar, dividir entre otras distribuciones

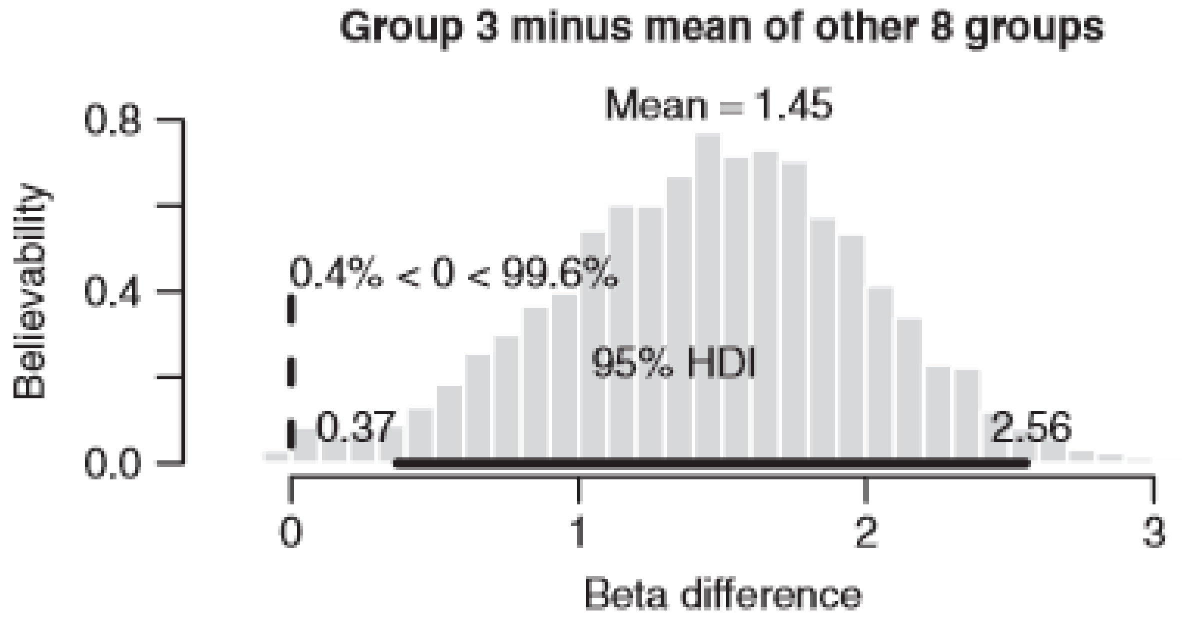


FIGURE 6 | Posterior estimate for difference of groups in Solari et al.[36] The third group is credibly different from the mean of the other eight groups.

HDI Bayesianos

El HDI de 95 % son los valores de Θ que tienen al menos ese nivel mínimo de credibilidad posterior.

- A diferencia del IC
  - El HDI tiene una interpretación directa en términos de la credibilidad de los valores de Θ
  - Es explícitamente acerca de $p(\Theta|D)$, que es exactamente de lo que queremos hablar
  - El HDI no depende del diseño de la investigación (el número de pruebas)
  - El HDI responde a las creencias a priori de las investigadoras, como debe ser.

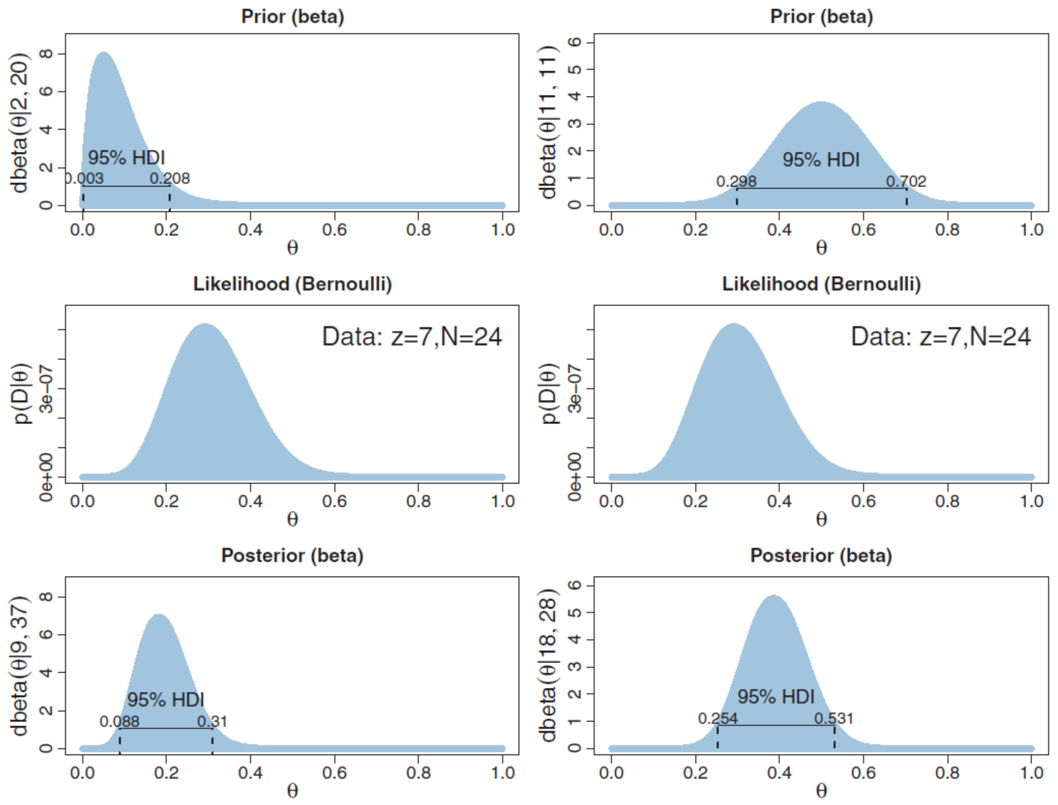


Figure 11.7 Posterior HDI for the bias of a Bernoulli process, when the prior assumes a tail-strong nail (left column) or a fair coin (right column).

Bayes vs frecuencias

Tenemos distribuciones completas de los valores de los parámetros

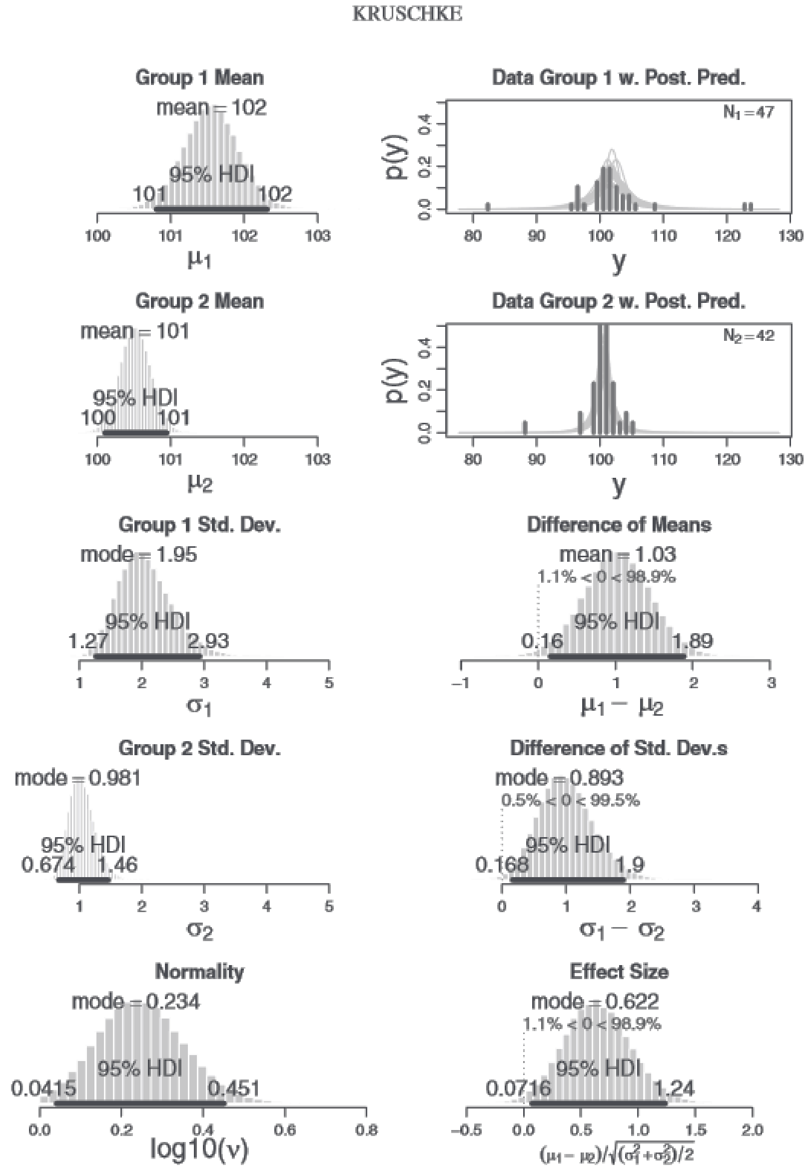
- No solo un estimador puntual

Tenemos precision explícita de los parámetros de interés

- No afectadas por las intenciones de muestreo y de prueba

Modelo flexible que es robusto a valores extremos

- No hay supuesto de normalidad



Evaluación de hipótesis nula en Bayes

¿Qué hacer si quieres evaluar la probabilidad de un valor nulo?

Dos caminos:

1. **Estimación vía parámetros: Región de Equivalencia Práctica (ROPE) y el Intervalo de más alta densidad (HDI)**

2. Comparación bayesiana de modelos: Decisión basada en factores bayesianos (Bayes factors)

1. Estimación de parámetros

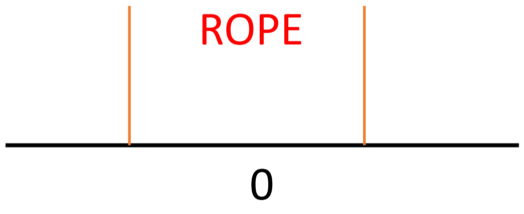
Usa la distribución posterior para discernir los valores creíbles del parámetro.

Considere un valor de referencia a evaluar del parámetro de interés: 0, -.1 - .1, 100 – 200.

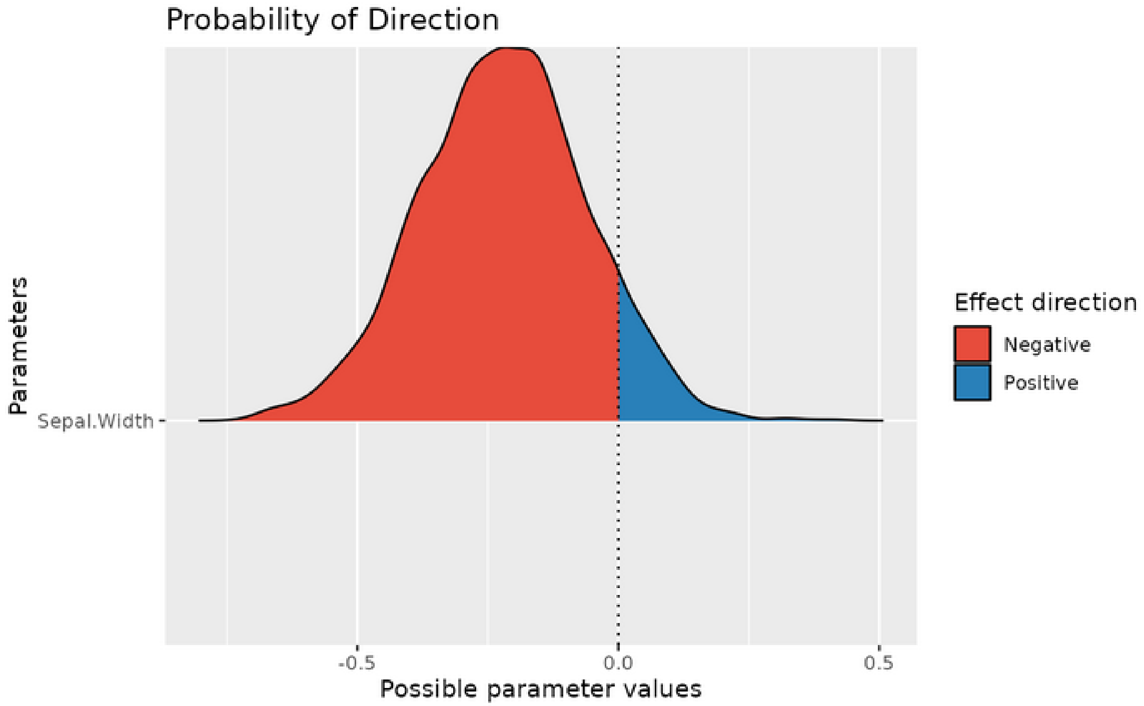
- Región de equivalencia en la práctica (ROPE): rango (del parámetro) considerado como el mismo en términos prácticos.

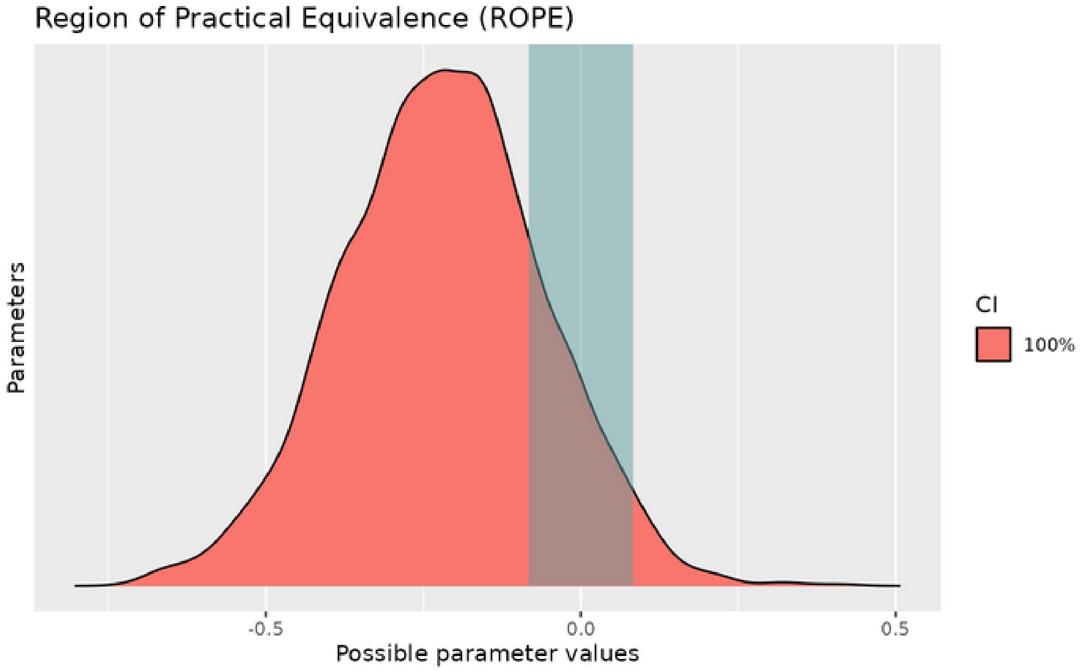
- Pero, ¿cómo se determina la ROPE?

    - Estadística vs Teóricamente significativo



Region of practical equivalence (ROPE)

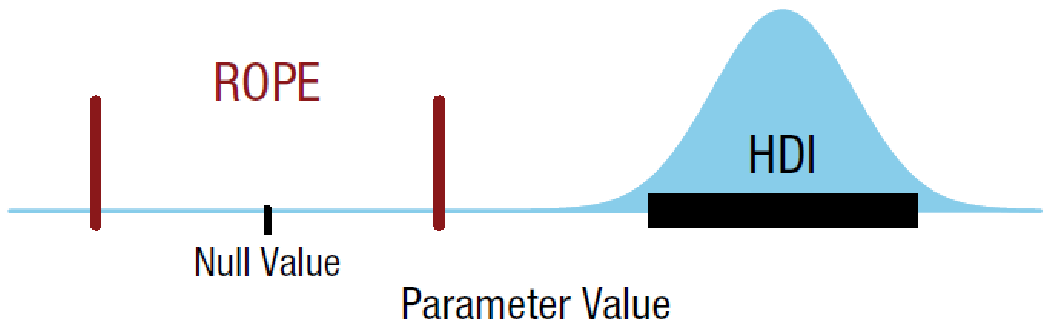





```
percentage_in_rope
```

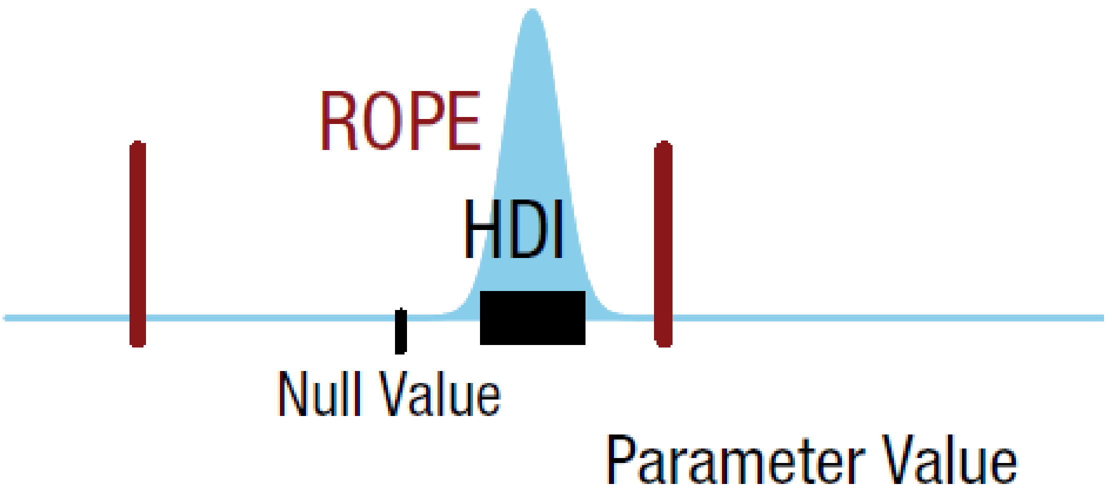
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado como **no creíble**, o rechazado, si su ROPE yace fuera del HDI relevante (típicamente el de 95 %) de su distribución posterior.



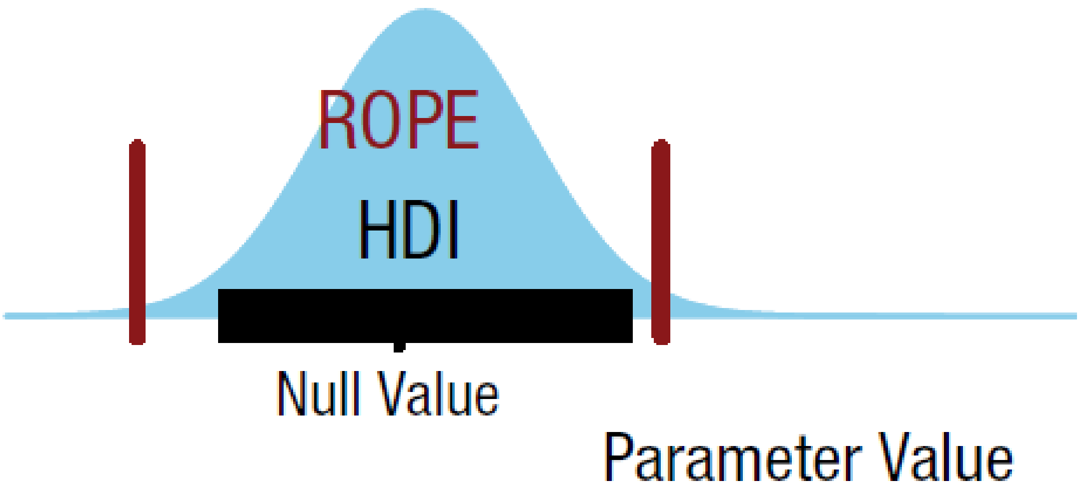
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado **aceptado** si su ROPE contiene completamente el HDI relevante (típicamente el de 95 %) de su distribución posterior.



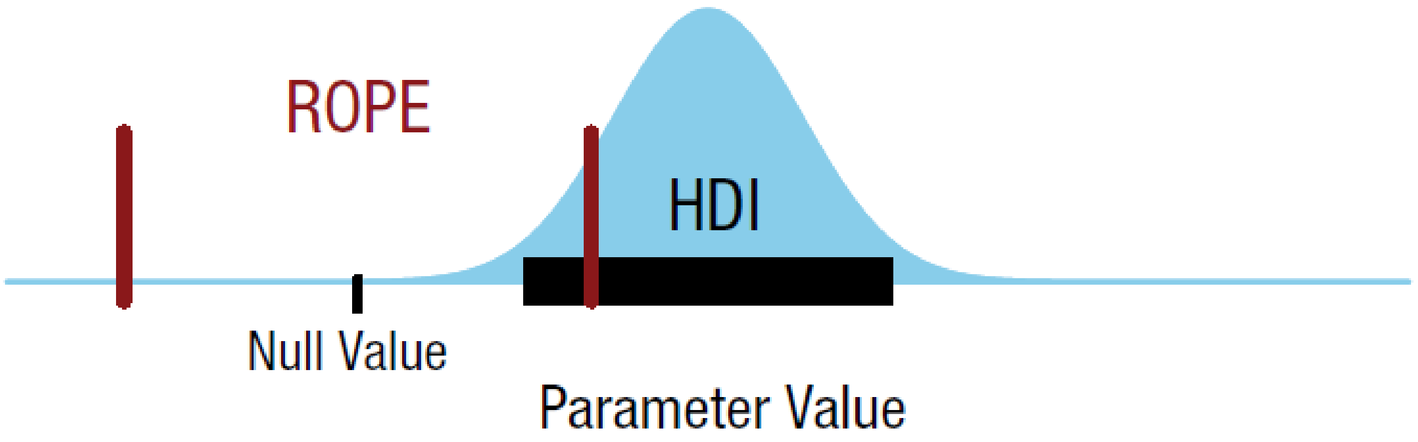
Estimación e inferencia a partir de distribuciones de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado **aceptado** si su ROPE contiene completamente el HDI relevante (típicamente el de 95 %) de su distribución posterior.



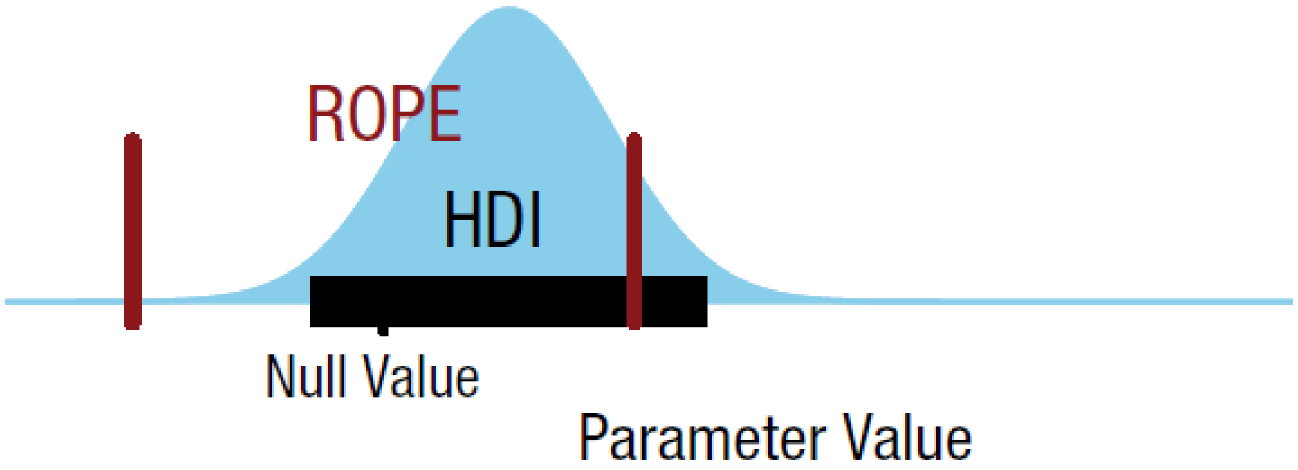
Estimación de parámetros

- Regla de decisión
  - ¿Qué hacer si el HDI y la ROPE traslapan, sin que la ROPE cubra la HDI por completo (ninguna de las condiciones anteriores se satisface?



Estimación de parámetros

- Regla de decisión
  - ¿Qué hacer si el HDI y la ROPE traslapan, sin que la ROPE cubra la HDI por completo (ninguna de las condiciones anteriores se satisface?



No siempre el 95% es simétrico

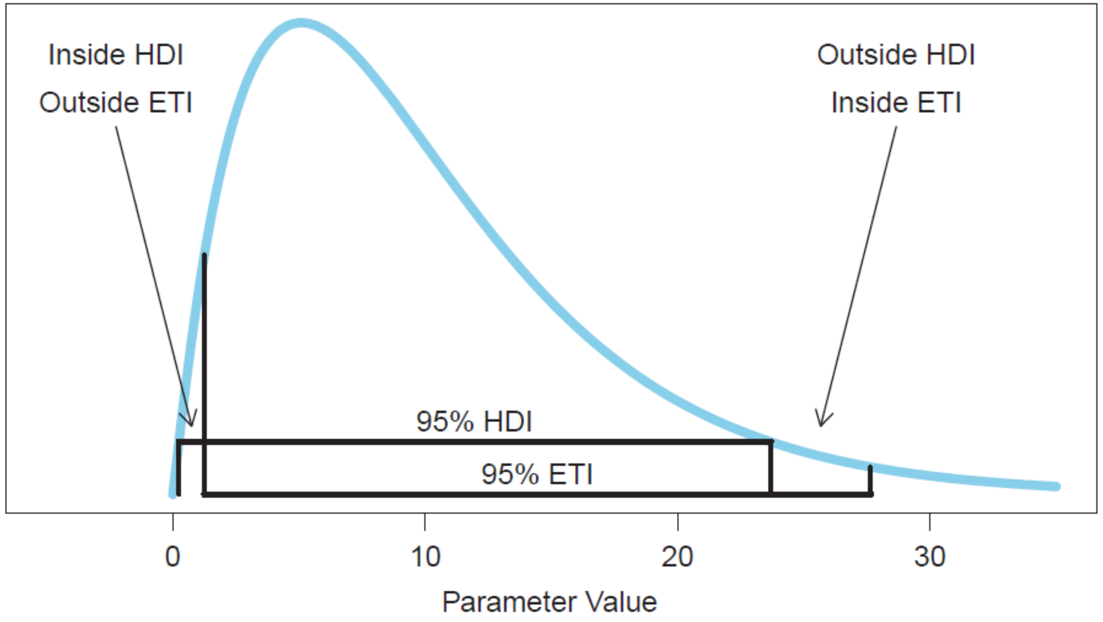
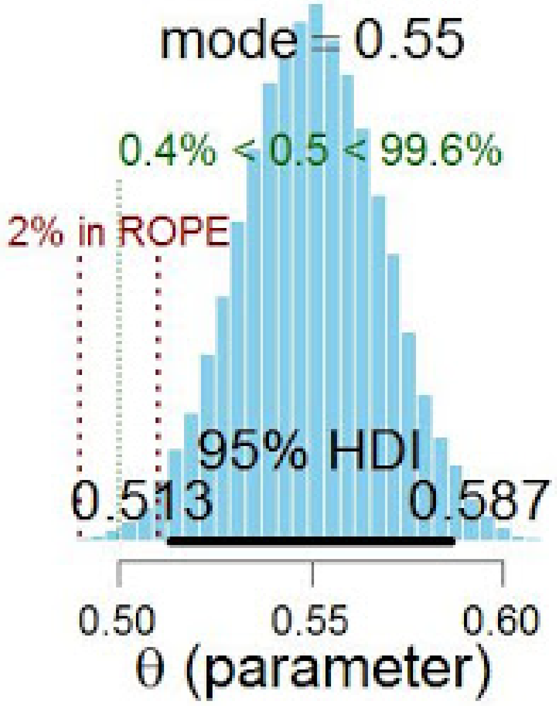


Figure 12.2: A skewed distribution has different 95% highest density interval (HDI) than 95% equal-tailed interval (ETI). Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

- Esto es importante para las desviaciones estándar
- Pero también para cierto tipo de parámetros en modelos espaciales o en la estimación de valores atípicos

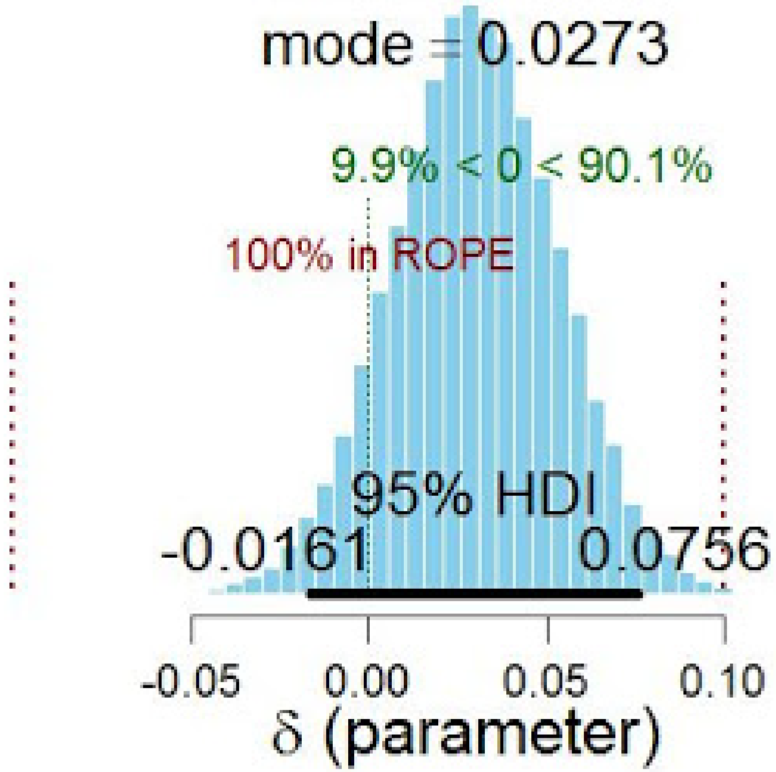
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado como **no creíble**, o rechazado, si su ROPE yace fuera del HDI relevante (típicamente el de 95 %) de su distribución posterior.



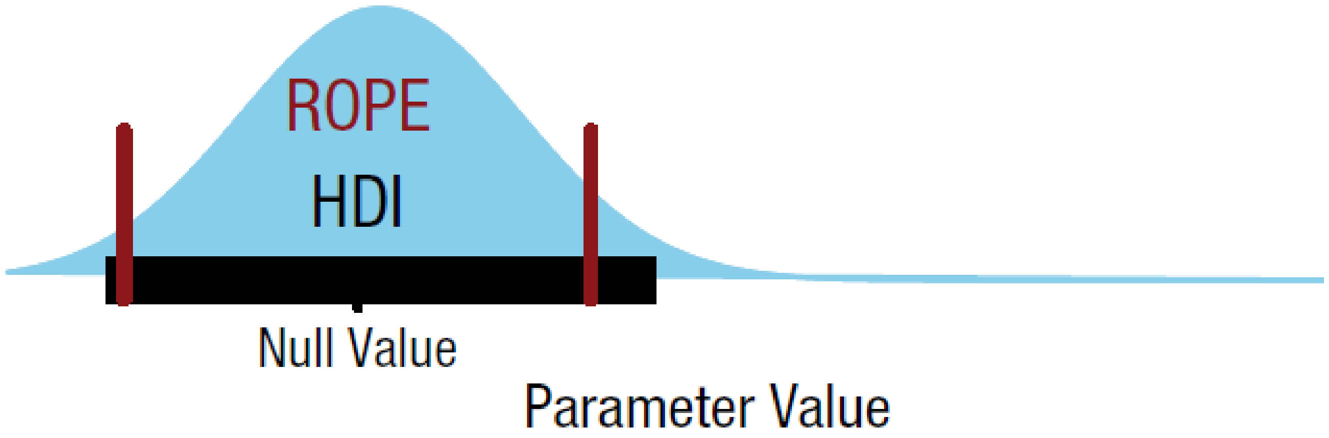
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado **aceptado** si su ROPE contiene completamente el HDI relevante (típicamente el de 95 %) de su distribución posterior.



Estimación de parámetros

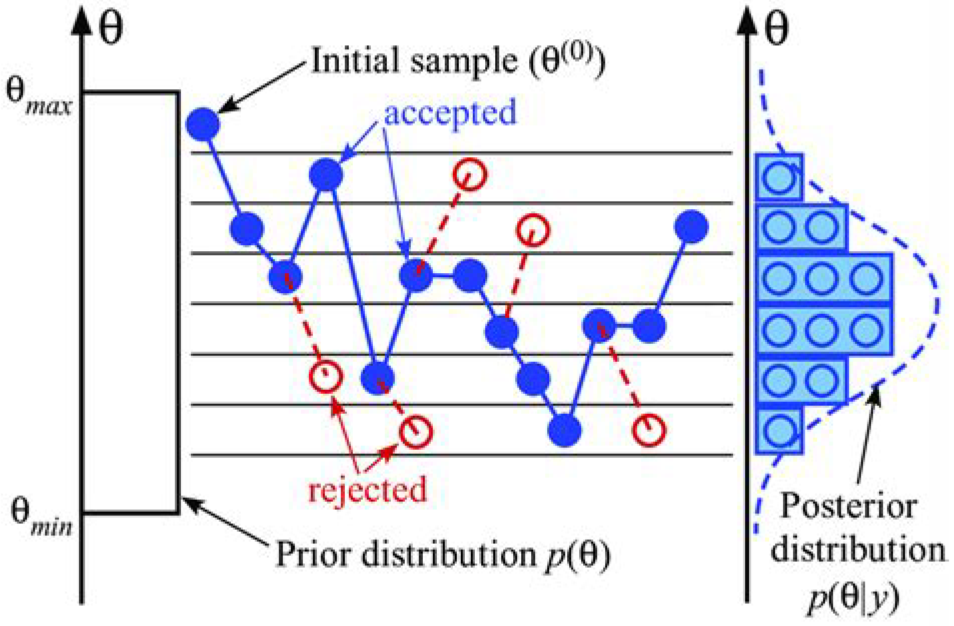
- Regla de decisión
  - ¿Qué hacer si el HDI y la ROPE traslapan, sin que la ROPE cubra la HDI por completo (ninguna de las condiciones anteriores se satisface?



Estimación de parámetros

- ¿Cómo se obtienen los parámetros?

- No olvidar la inestabilidad de las fronteras de las HDI debido a la aleatoriedad del MCMC
- Es importante advertir que la regla de decisión es independiente de la inferencia bayesiana (la parte bayesiana se acaba en la estimación de la distribución posterior)



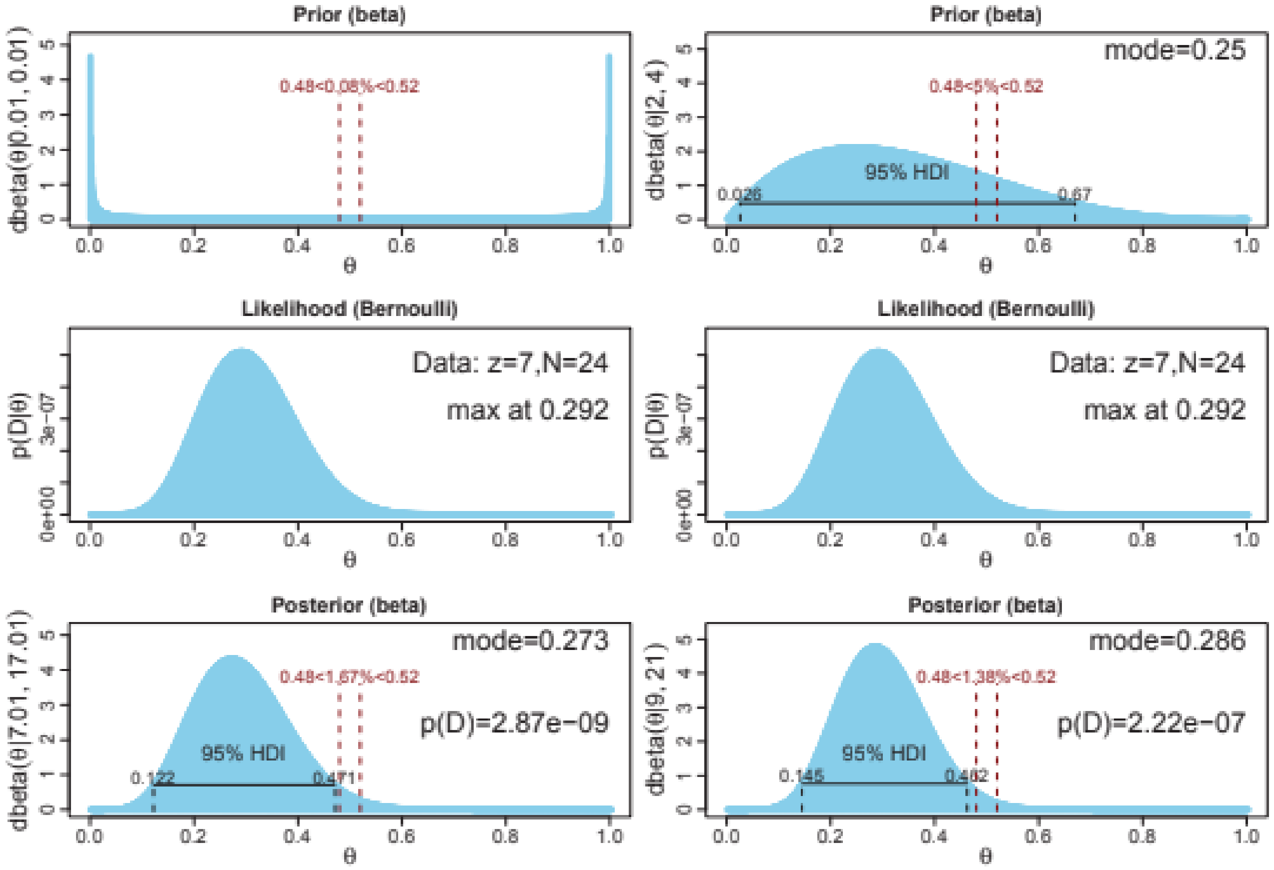
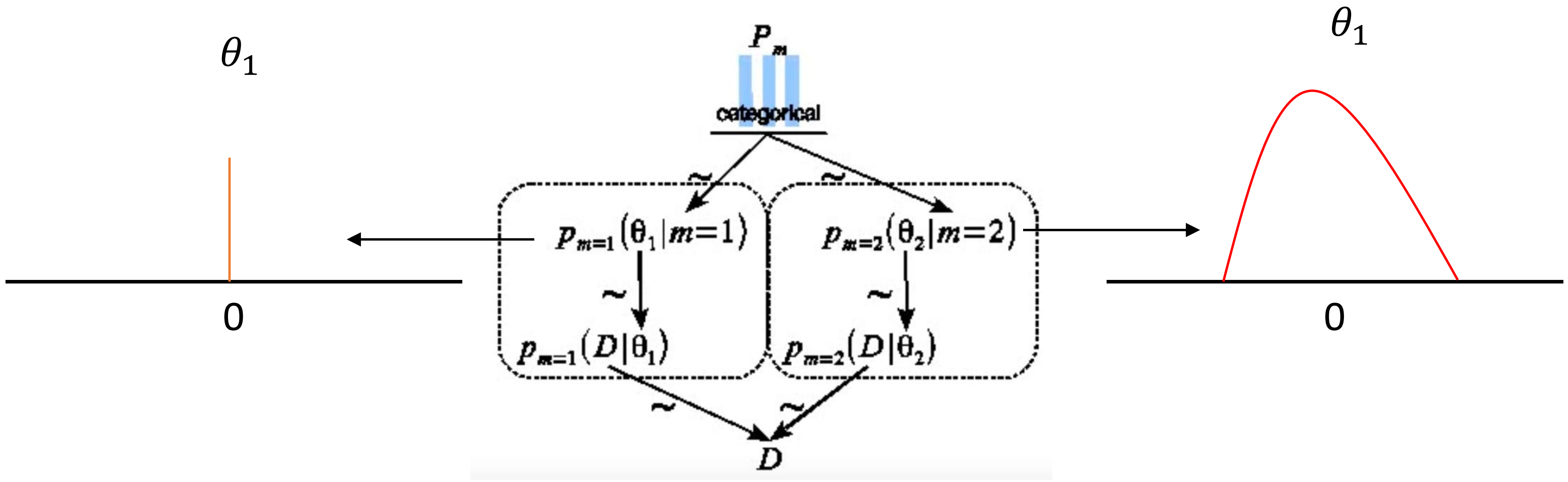


Figure 12.3: Left column: Haldane prior. Right column: Mildly informed prior. Vertical dashed lines mark a ROPE from 0.48 to 0.52. Annotation above the dashed lines indicates the percentage of the distribution within the ROPE. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

2. Estimación de la nula vía modelos

- Planteamos la pregunta:
  - Estimación puntual del valor nulo v distribución de posibilidades



Próxima clase

- Cómputo bayesiano:}

- Cómo se obtienen las distribuciones posteriores

Referencias

- Gelman, A., and Loken, E. (2014), "The Statistical Crisis in Science," American Scientist, 102, 460–465.