

Lecture 4: Intervals

Professor Alexander Franks

2025-10-13

Uncertainty?

Announcements

- Reading: Chapter 8.1 (intervals), 8.3 (posterior prediction)
- Homework 2 out today
- Quiz |, today.

Reminder: Frequentist confidence interval

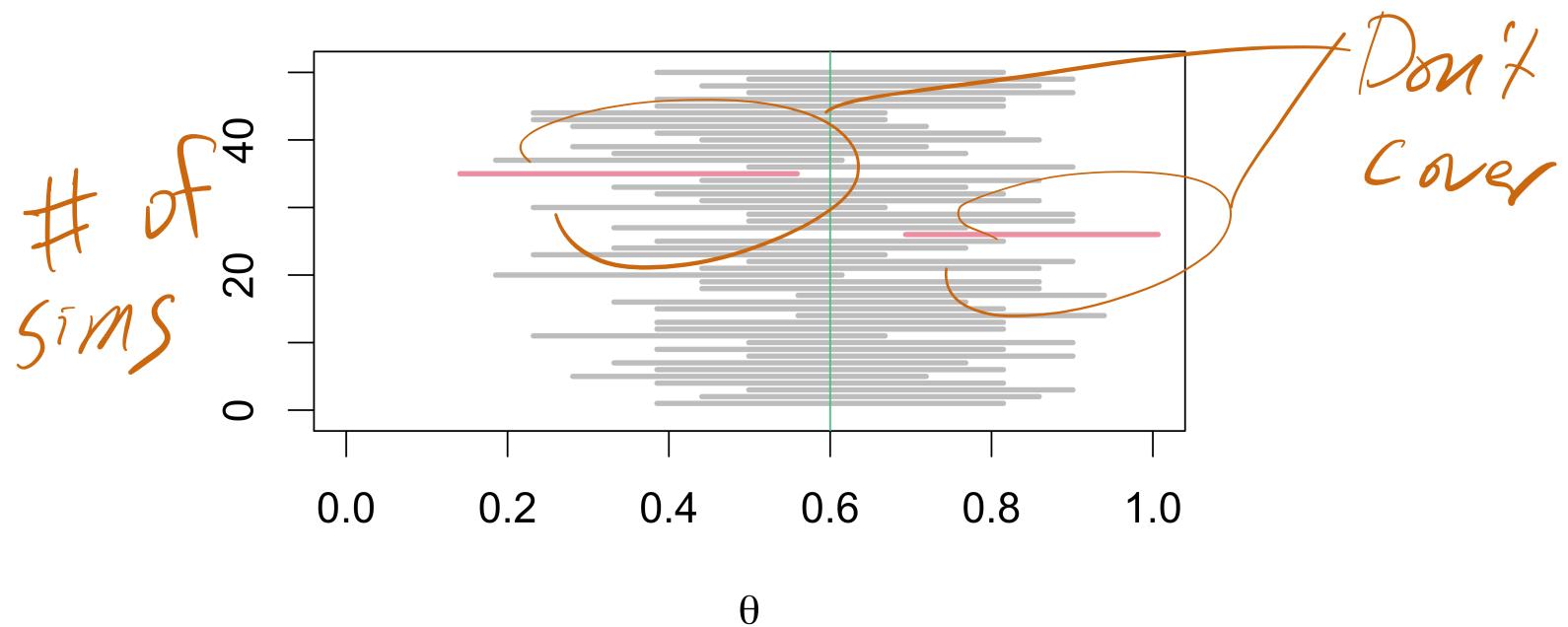
Non-Bayesian

- Frequentist interval: $Pr(l(Y) < \theta < u(Y) | \theta) = 0.95$
 - Probability that the interval will cover the true value *before* the data are observed.
 - Interval is random since Y is random

Fixed,

Random

Reminder: Frequentist confidence interval



We expect $0.05 \times 50 = 2.5$ will *not* cover the true parameter 0.6

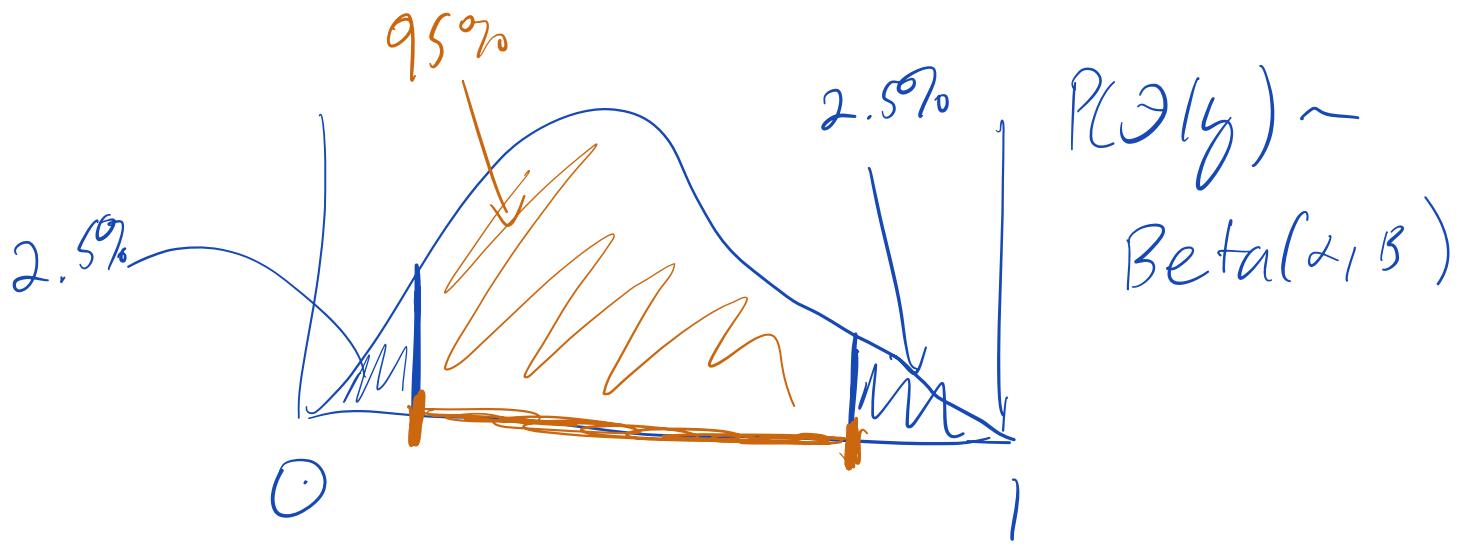
Posterior Credible Intervals

- Frequentist interval: $Pr(l(Y) < \theta < u(Y) | \theta) = 0.95$
 - Probability that the interval will cover the true value *before* the data are observed.
 - Interval is random since Y is random
- Bayesian Interval: $Pr(l(y) < \theta < u(y) | Y = y) = 0.95$
 - Information about the true value of θ *after* observing $Y = y$.
 - θ is random (because we include a prior), y is observed so interval is non-random.

Random.

Constant

Conditional
on Data



Quantile Based Interval

$$l(y) \leftarrow qbeta(.025, \alpha, \beta) \quad \text{Fixed Quantiles}$$

$$u(y) \leftarrow qbeta(1 - .025, \alpha, \beta)$$

Posterior Credible Intervals (Quantile-based)

- The easiest way to obtain a confidence interval is to use the quantiles of the posterior distribution.

If we want $100 \times (1 - \alpha)$ interval, we find numbers $\theta_{\alpha/2}$ and $\theta_{1-\alpha/2}$ such that:

$$1. p(\theta < \theta_{\alpha/2} \mid Y = y) = \alpha/2 \quad \text{lower} \quad \begin{matrix} \text{e.g. } \alpha = 5\% \\ = .05 \end{matrix}$$

$$2. p(\theta > \theta_{1-\alpha/2} \mid Y = y) = \alpha/2 \quad \text{upper}$$

$$\text{U}(y) \quad \text{L}(y), \text{U}(y)$$

$$p(\theta \in [\theta_{\alpha/2}, \theta_{1-\alpha/2}] \mid Y = y) = 1 - \alpha$$

- Use quantile functions in R, e.g. `qbeta`, ~~`qpois`~~, `qnorm` etc.

ggamma,

Example: interval for shooting skill

- The posterior distribution for Covington's shooting percentage is a

$$\text{Beta}(49 + 478, 50 + 873) = \text{Beta}(528, 924)$$

- For a 95% *credible* interval, $\alpha = 0.05$
 - Lower endpoint: `qbeta(0.025, 528, 924)`
 - Upper endpoint: `qbeta(0.975, 528, 924)`
 - $[\theta_{\alpha/2}, \theta_{1-\alpha/2}] = [0.34, 0.39]$
- Compared to frequentist *confidence* interval without prior information: [0.39, 0.59]
- End-of-season percentage was 0.37
- Credible intervals and confidence intervals have different

Highest Posterior Density (HPD) region

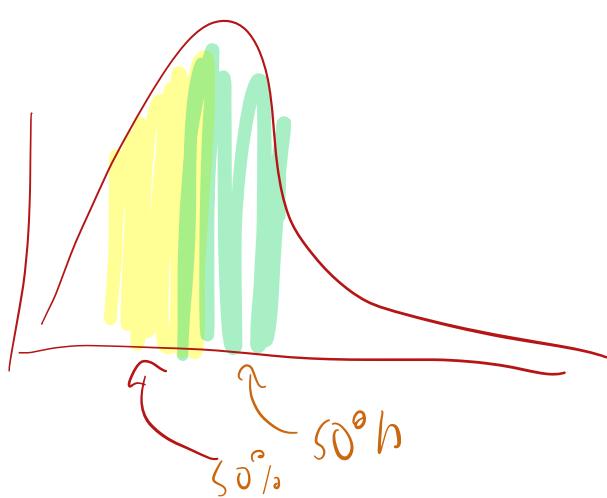
Definition: (HPD region) A $100 \times (1 - \alpha)$ HPD region consists of a subset of the parameter space, $R(y) \in \Theta$ such that

1. $\Pr(\theta \in R(y)|Y = y) = 1 - \alpha = 95\%$

- The probability that θ is in the HPD region is $1 - \alpha$

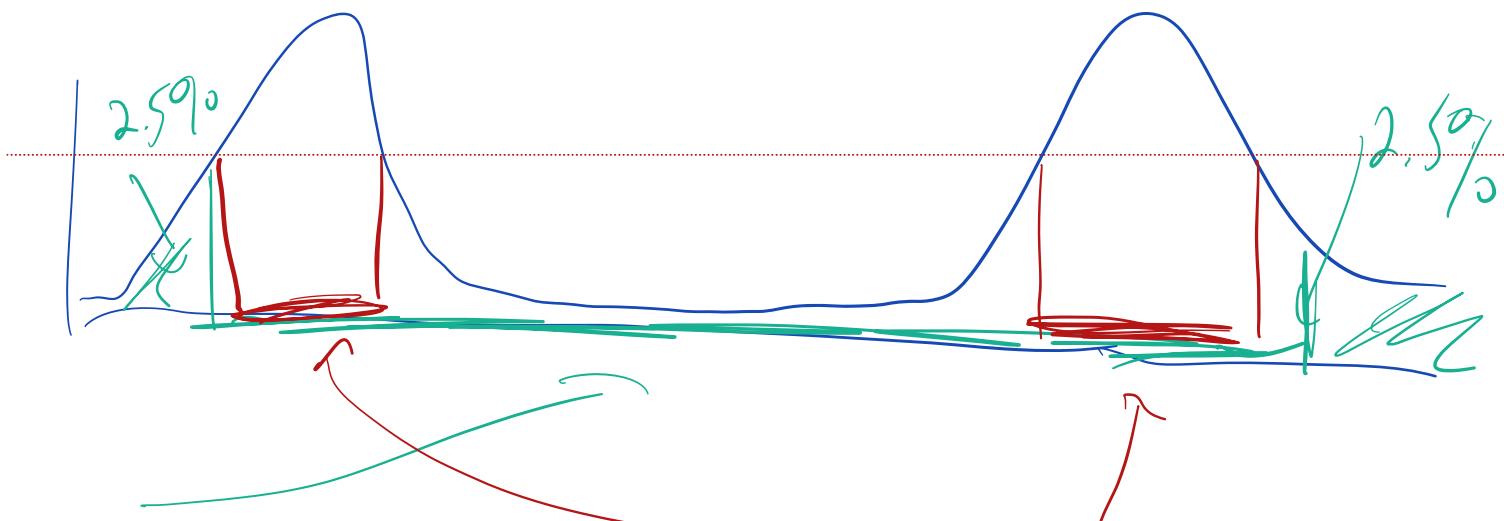
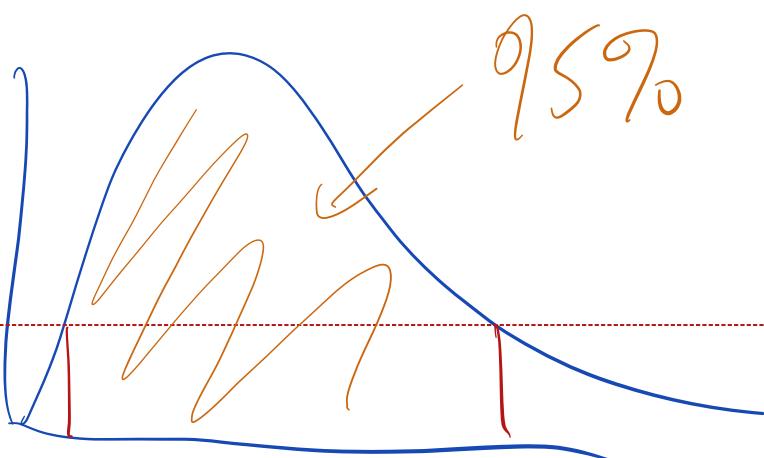
2. If $\theta_a \in R(y)$, and $\theta_b \notin R(y)$ then $p(\theta_a|Y = y) > p(\theta_b|Y = y)$] Math

- All points in an HPD region have a higher posterior density than points outside the region.



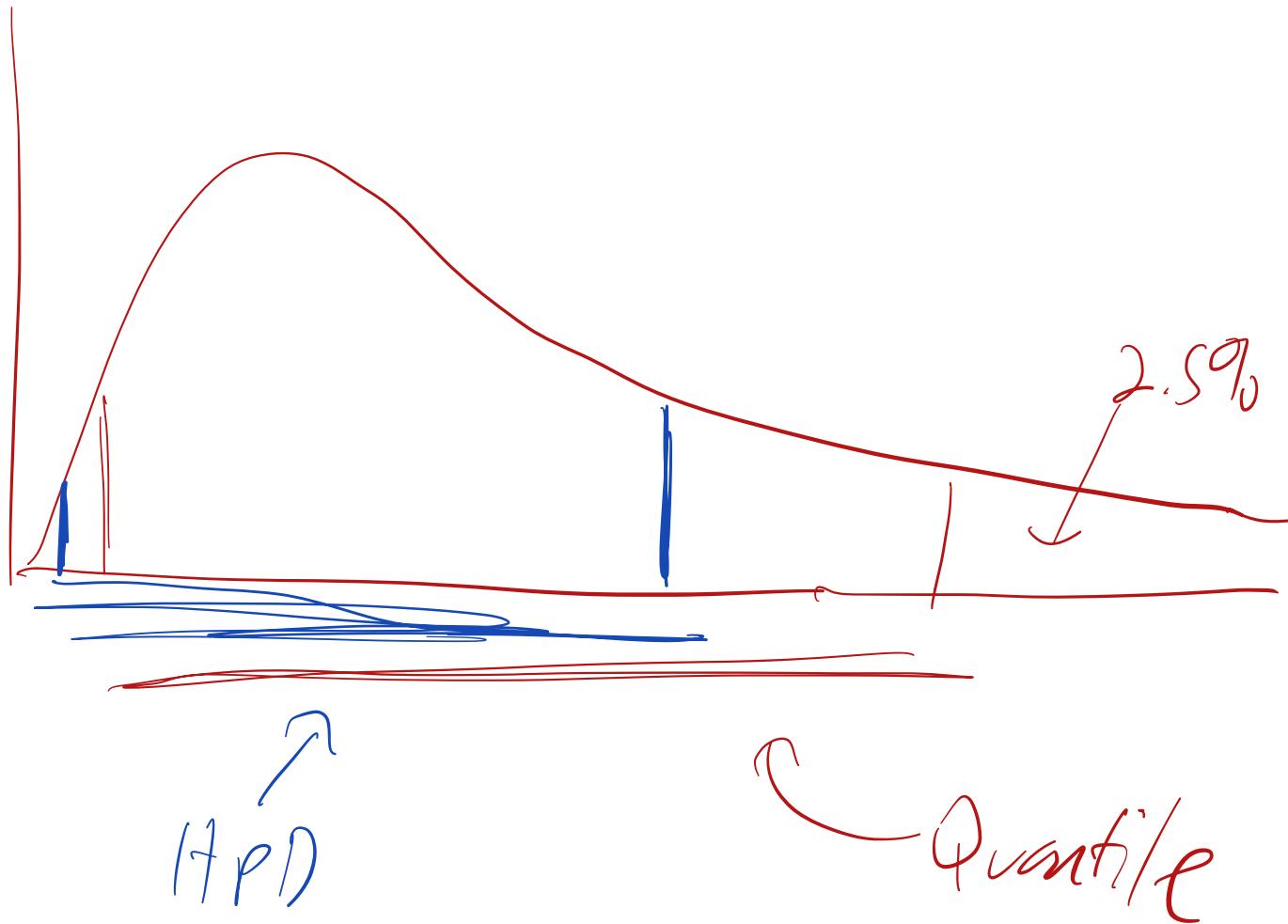
HPD interval
(region) is
the shortest
possible

$1-\alpha\%$ interval
possible.



95% Quantile Based.

95% HPD -



Highest Posterior Density (HPD) region

95%

1. $\underline{p(\theta \in s(y) \mid Y = y)} = 1 - \alpha$
 2. If $\theta_a \in s(y)$, and $\theta_b \notin s(y)$, then
 $p(\theta_a \mid Y = y) > p(\theta_b \mid Y = y).$
- All points in an HPD region have a higher posterior density than points out-side the region.

The HPD region is the *smallest* region with prob $(1 - \alpha)\%$

Calibration: Frequentist Behavior of Bayesian Intervals

Match Reality.

- A credible interval is calibrated if it has the right frequentist coverage
- Bayesian credible intervals usually won't have correct frequentist coverage
- If our prior was well-calibrated and the sampling model was correct, we'd have well-calibrated credible intervals
- Specifying nearly calibrated prior distributions is hard!

Calibration of political predictions

The best test of a probabilistic forecast is whether it's **well calibrated**. By that I mean: Out of all FiveThirtyEight forecasts that give candidates about a 75 percent shot of winning, do the candidates in fact win about 75 percent of the time over the long run? It's a problem if these candidates win only 55 percent of the time. But from a statistical standpoint, it's just as much of a problem if they win 95 percent of the time.

source: fivethirtyeight.com

Calibration of political predictions

Calibration for FiveThirtyEight “polls-plus” forecast

WIN PROBABILITY RANGE	NO. FORECASTS	EXPECTED NO. WINNERS	ACTUAL NO. WINNERS
95-100%	27	26.7	26
75-94%	15	13.1	14
50-74%	14	8.7	11
25-49%	13	4.8	3
5-24%	27	3.1	1
0-4%	88	0.8	1

source: <https://fivethirtyeight.com/features/when-we-say-70-percent-it-really-means-70-percent/>

The age guessing game*



[37, 55]



[37, 42]



[39, 41]

*Bayesian edition

Posterior Predictive Distributions

Posterior predictive distribution

- An important feature of Bayesian inference is the existence of a predictive distribution for new observations.
 - Let \tilde{y} be a new (unseen) observation, and y_1, \dots, y_n the observed data.
 - The Posterior predictive distribution is $p(\tilde{y} | y_1, \dots, y_n)$
- The predictive distribution does not depend on unknown parameters
- The predictive distribution only depends on observed data
- Asks: what is the probability distribution for new data given old data?

No Parameters

Another Basketball Example

- I take free throw shots and make 1 out of 2. How many do you think I will make if I take 10 more?
- If my true “skill” was 50%, then $\tilde{Y} \sim \text{Bin}(10, 0.50)$
- Is this the correct way to calculate the predictive distribution?

No. Don't know true θ !

Posterior Prediction

If you know θ , then we know the distribution over future attempts:

$$\tilde{Y} \sim \text{Bin}(10, \theta)$$

Posterior Prediction

- We already observed 1 make out of 2 tries.
- Assume a $\text{Beta}(1, 3)$ prior distribution
 - e.g. a priori you think I'm more likely to make 25% of my shots
- Then $p(\theta | Y = 1, n = 2)$ is a $\text{Beta}(2, 4)$
- Intuition: weight $\tilde{Y} \sim \text{Bin}(10, \theta)$ by $p(\theta | Y = 1, n = 2)$

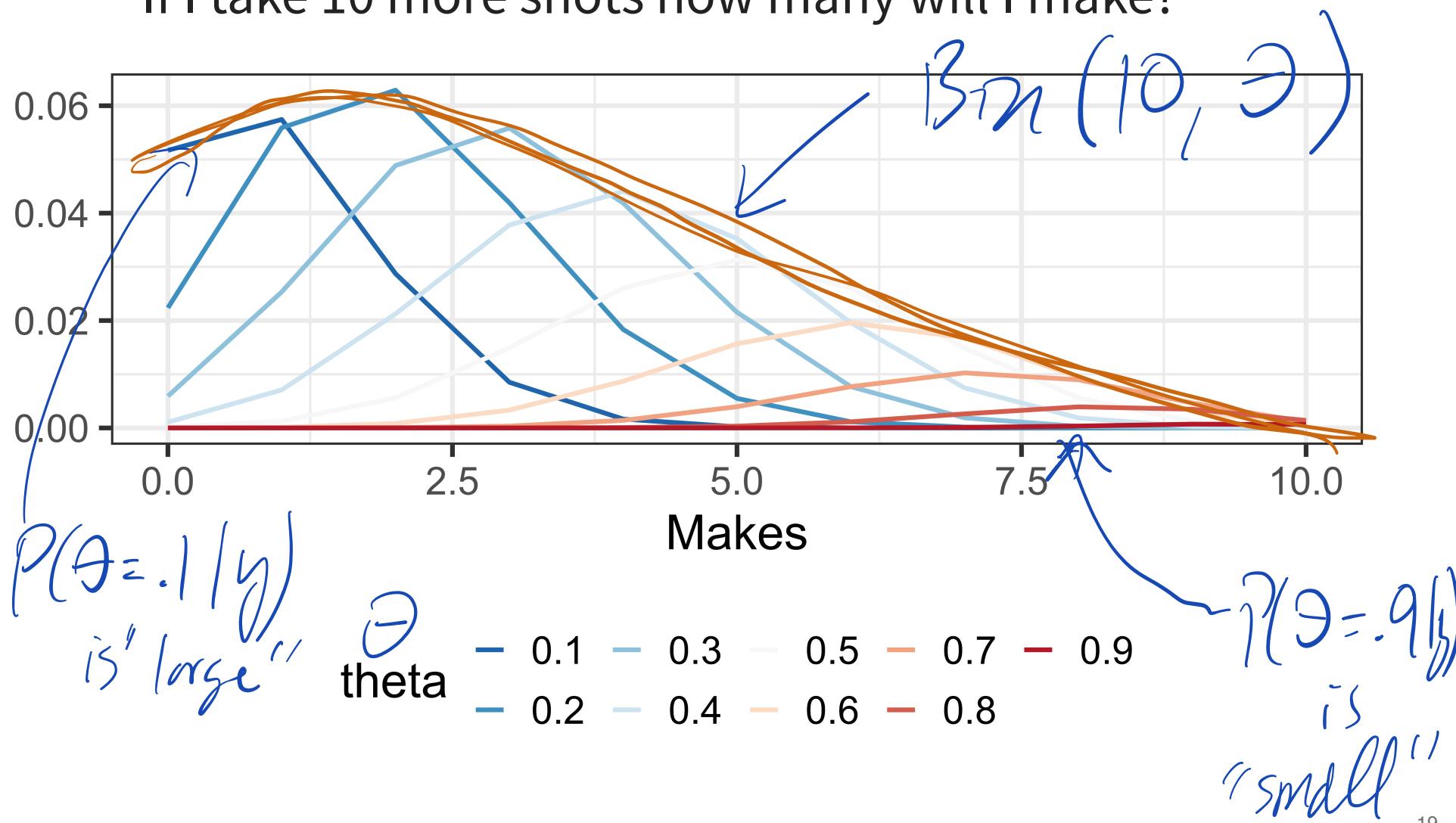
$$P(\theta | y) \sim \text{Beta}(\alpha + y, \beta + n - y)$$

Belief about
true skill.

Sampling
variability
(chance)

Posterior Prediction

If I take 10 more shots how many will I make?



Posterior predictive distribution

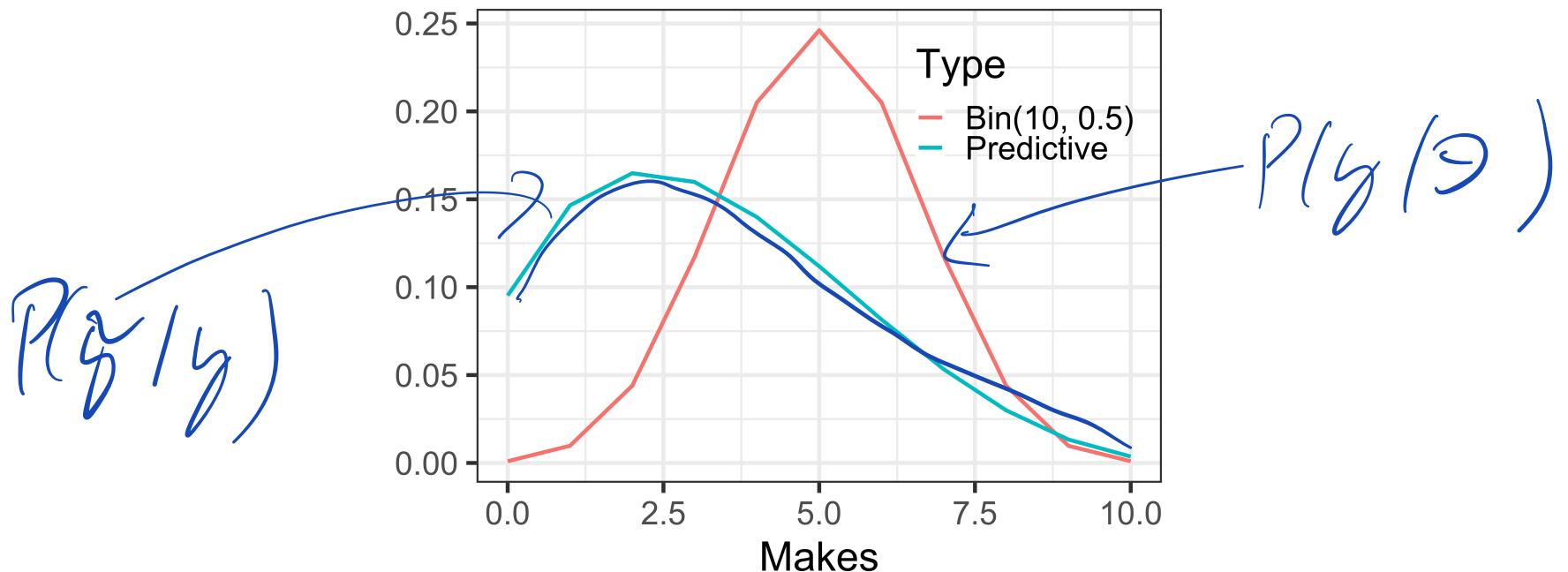
$$P(\tilde{y} | y_1, \dots, y_n) = \int P(\tilde{y} | \theta) P(\theta | y_1, \dots, y_n) d\theta$$

Bin if
 θ known.

Posterior
gives "weight"
associated
w/ θ
(e.g. Beta)

Posterior predictive distribution

$$p(\theta) = \text{Beta}(1, 3), p(\theta | y) = \text{Beta}(2, 4)$$



The predictive density, $p(\tilde{y} | y)$, answers the question “if I take 10 more shots how many will I make, given that I already made

The posterior predictive distribution

$$\begin{aligned} p(\tilde{y} \mid y_1, \dots, y_n) &= \int p(\tilde{y}, \theta \mid y_1, \dots, y_n) d\theta \\ &= \int p(\tilde{y} \mid \theta) p(\theta \mid y_1, \dots, y_n) d\theta \end{aligned}$$

- The posterior predictive distribution describes our uncertainty about a new observation after seeing n observations
- It incorporates uncertainty due to the sampling in a model $p(\tilde{y} \mid \theta)$ and our posterior uncertainty about the data generating parameter, $p(\theta \mid y_1, \dots, y_n)$

$$P(\tilde{\gamma} | \theta) \sim \text{Bin}(10, \theta)$$

$$P(\theta | y) \sim \text{Beta}(2, 4)$$

$$\begin{aligned} P(\tilde{\gamma} | y) &= \underbrace{\int_0^1 P(\tilde{\gamma} | \theta) P(\theta | y) d\theta}_{\text{Beta}(\alpha, \beta)} \\ &= \int_0^1 \binom{10}{\tilde{\gamma}} \theta^{\tilde{\gamma}} (1-\theta)^{10-\tilde{\gamma}} \underbrace{\theta^{\alpha-1} (1-\theta)^{\beta-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\text{Beta}(\alpha, \beta)} d\theta \end{aligned}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \binom{10}{\tilde{\gamma}} \int_0^1 \theta^{\tilde{\gamma}+1} (1-\theta)^{13-\tilde{\gamma}} d\theta$$

looks like
a unnormalized Beta

Beta-Binomial Distribution.

$$P(\tilde{\gamma} | y) = \binom{10}{\tilde{\gamma}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\tilde{\gamma}+1)\Gamma(13-\tilde{\gamma})}{\Gamma(13+1)}$$

Homework 1

Subjective Bayesianism

- So far we have focused on defining priors using domain expertise

- “Subjective” Bayes

Motivates pseudo-counts
in conjugate priors

- Essentially what we have discussed so far
- Priors usually represent subjective judgements can't always be rigorously justified

- Alternative: “objective” Bayes

Objective Bayesianism

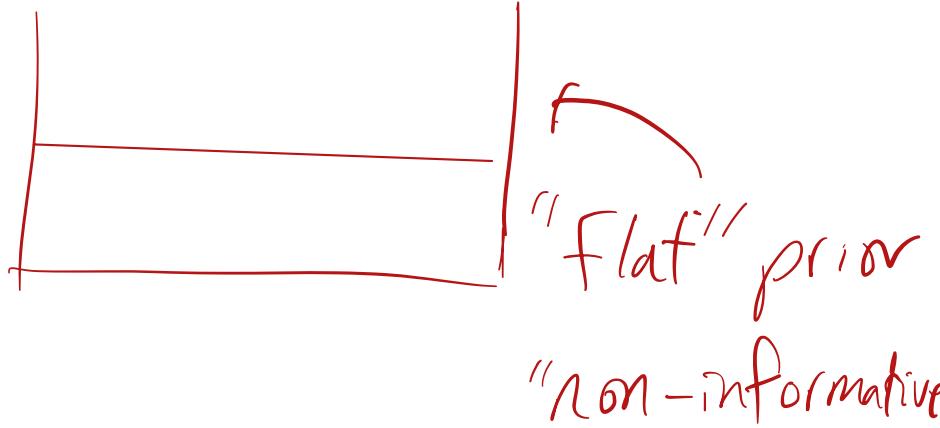
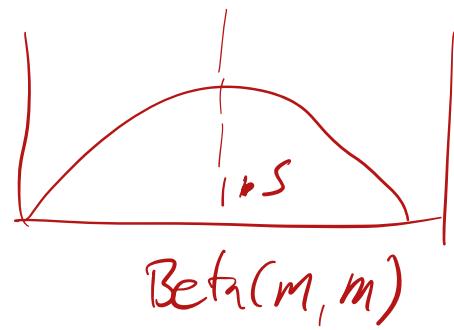
$P(\Theta)$

- Is there a way to define “objective” prior distributions?
 - Good default prior distributions for some problems?
 - “Non-informative” prior distributions? → “let the data speak”
- Also called “reference” or “default” priors
- Can we find prior distributions that lead to (approximately) correct frequentist calibration? *Freq matches Bayes?*
- Can we find prior distributions which minimize the amount of information contained in the distribution?
 - Principle of maximum entropy (MAXENT).

$$Y \sim \text{Bin}(n, \theta)$$

$$\theta \sim \text{Beta}(\alpha, \beta)$$

$$\alpha = \beta = 1$$



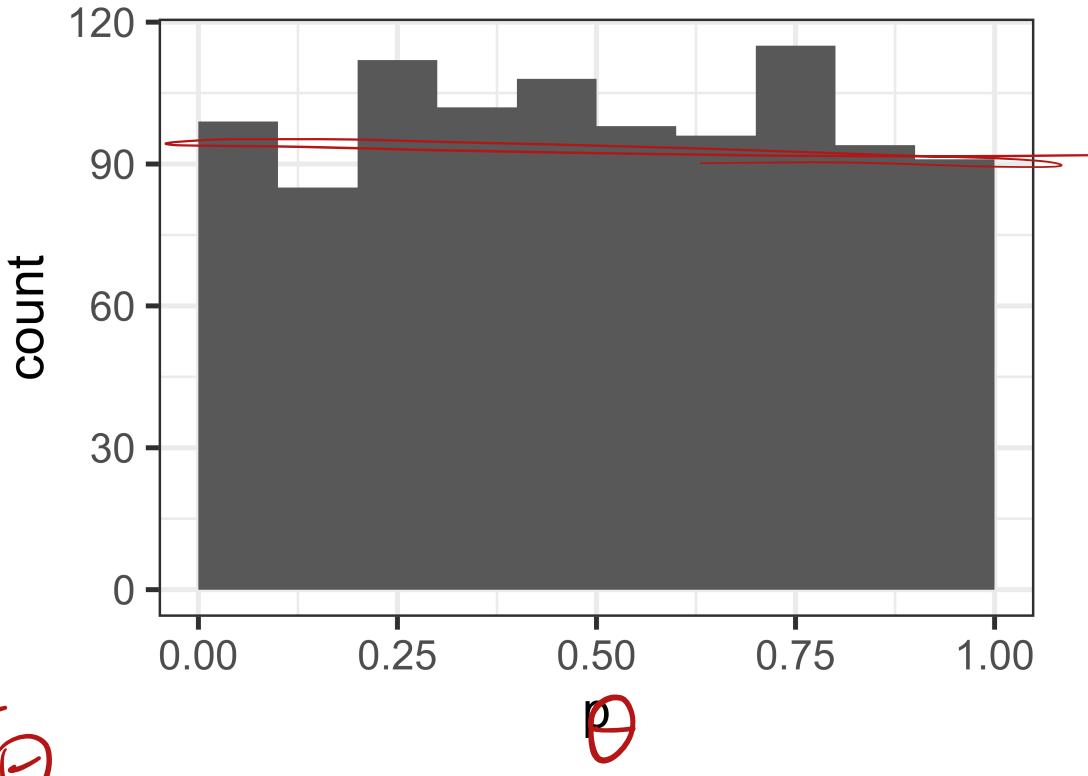
Difficulties with non-informative priors

Uniform distribution for p

$$P(\theta) \sim \text{Unif}$$

```
1 p <- runif(1000)
2 tibble(p=p) %>% ggplot() +
  geom_histogram(aes(x=p), boundary=0.5, binwidth=0.1) +
  theme_bw(base_size=24)
```

What if
I care
about the
odds



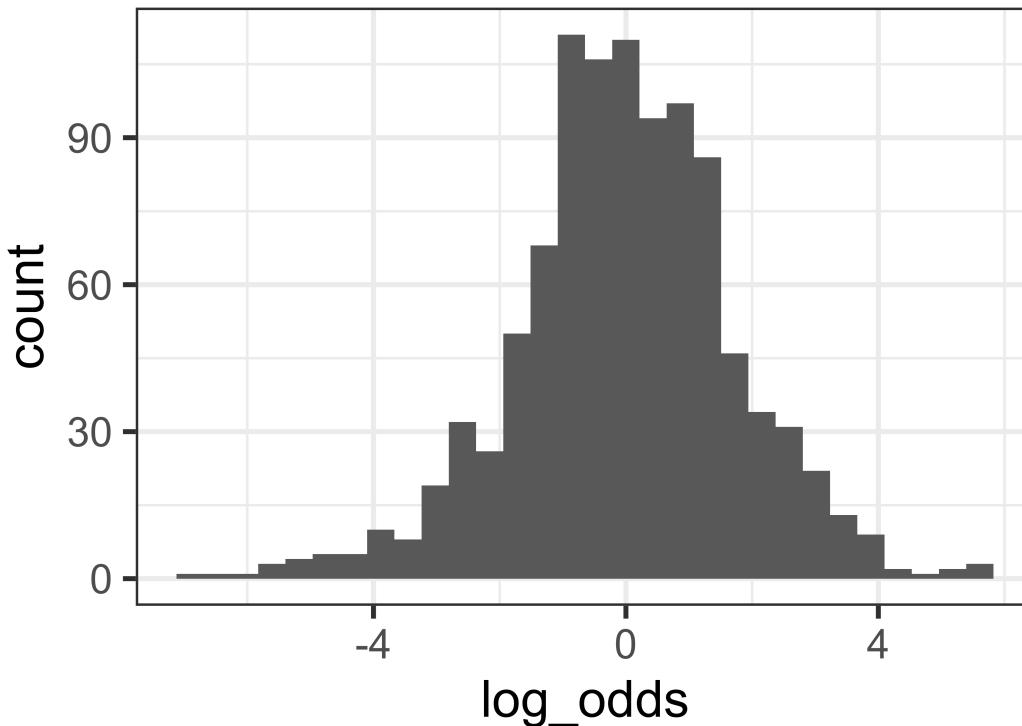
$$\text{Odds} = \frac{\theta}{1-\theta}$$

Difficulties with non-informative priors

Implied distribution for odds = $p/(1-p)$

```
1 log_odds <- log(p/(1-p))  
2 tibble(log_odds=log_odds) %>% ggplot() +  
3   geom_histogram(aes(x=log_odds)) +  
4   theme_bw(base_size=24)
```

Uniform on



θ ,
non-uniform
for log odds!!

$$X \sim \text{Pois}(\lambda), \quad \lambda \in \mathbb{R}^+ \quad (\text{positive real number})$$

$[0, \infty)$

$$P(\lambda) \sim \text{Unif}[0, \infty)$$

$$\int_0^\infty P(\lambda) d\lambda = \int_0^\infty 1 d\lambda = \infty$$

$$P(\lambda | y) \propto \underbrace{e^{-\lambda} \lambda^y}_{L(\lambda)} \underbrace{I[\lambda \geq 0]}_{P(\lambda)} \propto e^{-\lambda} \lambda^y$$

$$\int_0^\infty e^{-\lambda} \lambda^y < \infty$$

Improper prior distributions

- For the Beta distribution we chose a uniform prior, where $p(\theta) \propto \text{const}$. This was ok because:
 - $\int_0^1 p(\theta)d\theta = \text{const} < \infty$
 - We say this prior distribution is *proper* because it is integrable
- For the Poisson distribution, try the same thing:
 $p(\lambda) \propto \text{const}$
 - $\int_0^\infty p(\lambda)d\lambda = \infty$
 - In this case we say $p(\lambda)$ is an *improper prior*

Not a distribution!

Improper prior distributions

- Sometimes there is an absence of precise prior information
- The prior distribution does not have to be proper but the posterior does!
 - A proper distribution is one with an integrable density
 - If you use an improper prior distribution, you need to check that the posterior distribution is also proper

Must have Proper Posterior.

Summary

- Bayesian credible intervals
 - Posterior probability that the value falls in the interval
 - Still strive for well-calibrated intervals (in the frequentist sense)
- Posterior predictive distributions
 - Estimated distribution for new data our uncertainty about the parameters
- Non-informative prior distributions