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 $p(\tilde{y} \mid y_1, \dots y_n)$

$P(y|\theta)$ - Sampling. $\rightarrow$ likelihood, $L(\theta)$

$P(\theta)$ - prior

$P(\theta|y)$ - posterior: Lik × Prior

$P(\tilde{y}|y)$ - posterior predictive.

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 $y_1, \ldots y_n$ the observed data.
  - The Posterior predictive distribution is $p(\tilde{y} \mid y_1, \ldots y_n)$
- 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 observations of old data?

Another Basketball Example

$\hat{\theta}_{MLE} = 0.5$

- I take free throw shots and make 1 out of 2. How many do you think I will make if I take 10 more?

??

θ

- If my true "skill" was 50%, then $\tilde{Y} \sim \text{Bin}(10, 0.\cancel{50})$
- Is this the correct way to calculate the predictive distribution?

Predictive Data.

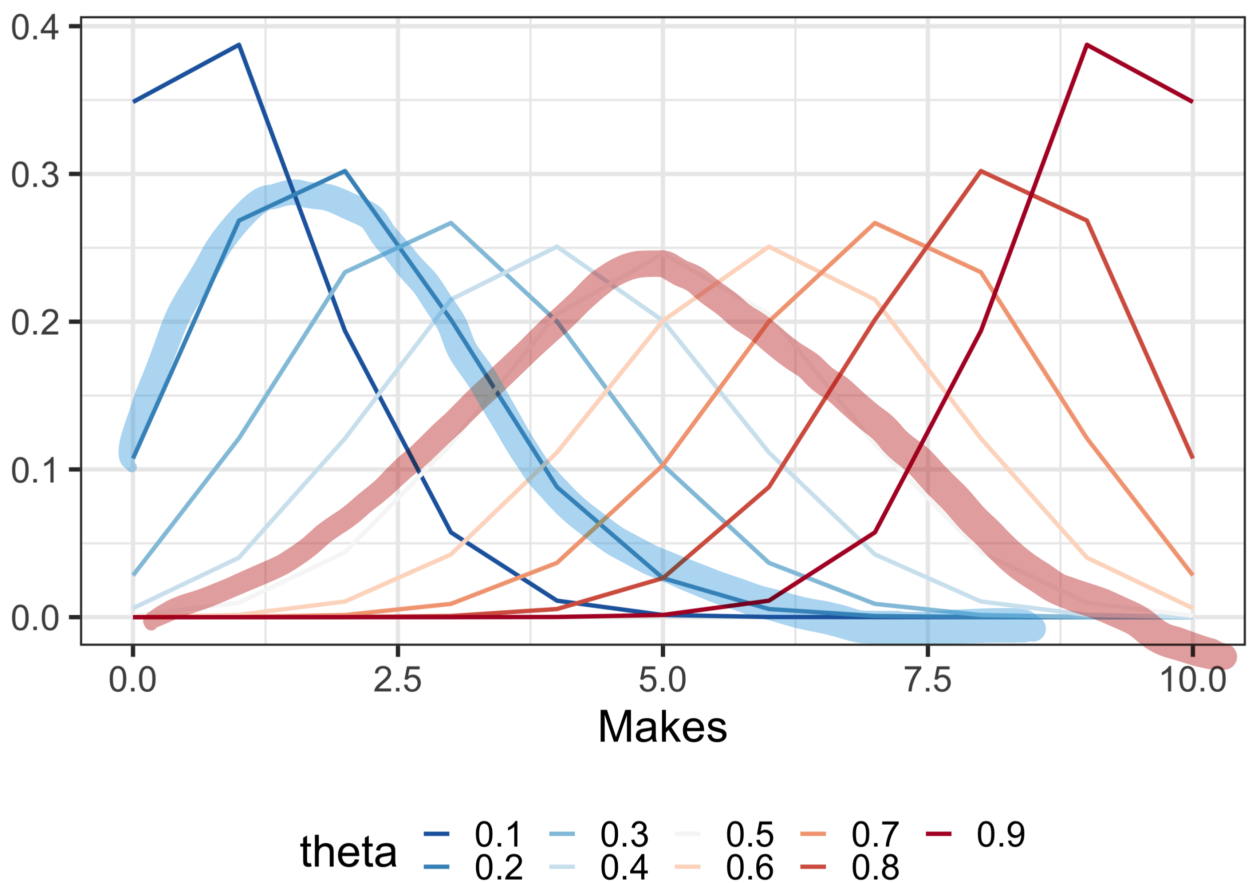
"True skill" θ, Fraction made w/ ∞ shots.

$$\int \underbrace{p(\tilde{y} \mid \theta)}_{\text{sampling variability}} \underbrace{p(\theta \mid y)}_{\text{uncertainty about } \theta} d\theta = \underline{p(\tilde{y} \mid y)}$$

Posterior Prediction

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

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



Posterior Prediction

$y \sim Bin(2, \theta)$

- We already observed 1 make out of 2 tries.

$y = 1$

- 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 $p(\theta \mid Y = 1, n = 2)$ is a $\text{Beta}(2, 4)$
- Intuition: weight $\tilde{Y} \sim \text{Bin}(10, \theta)$ by $p(\theta \mid Y = 1, n = 2)$

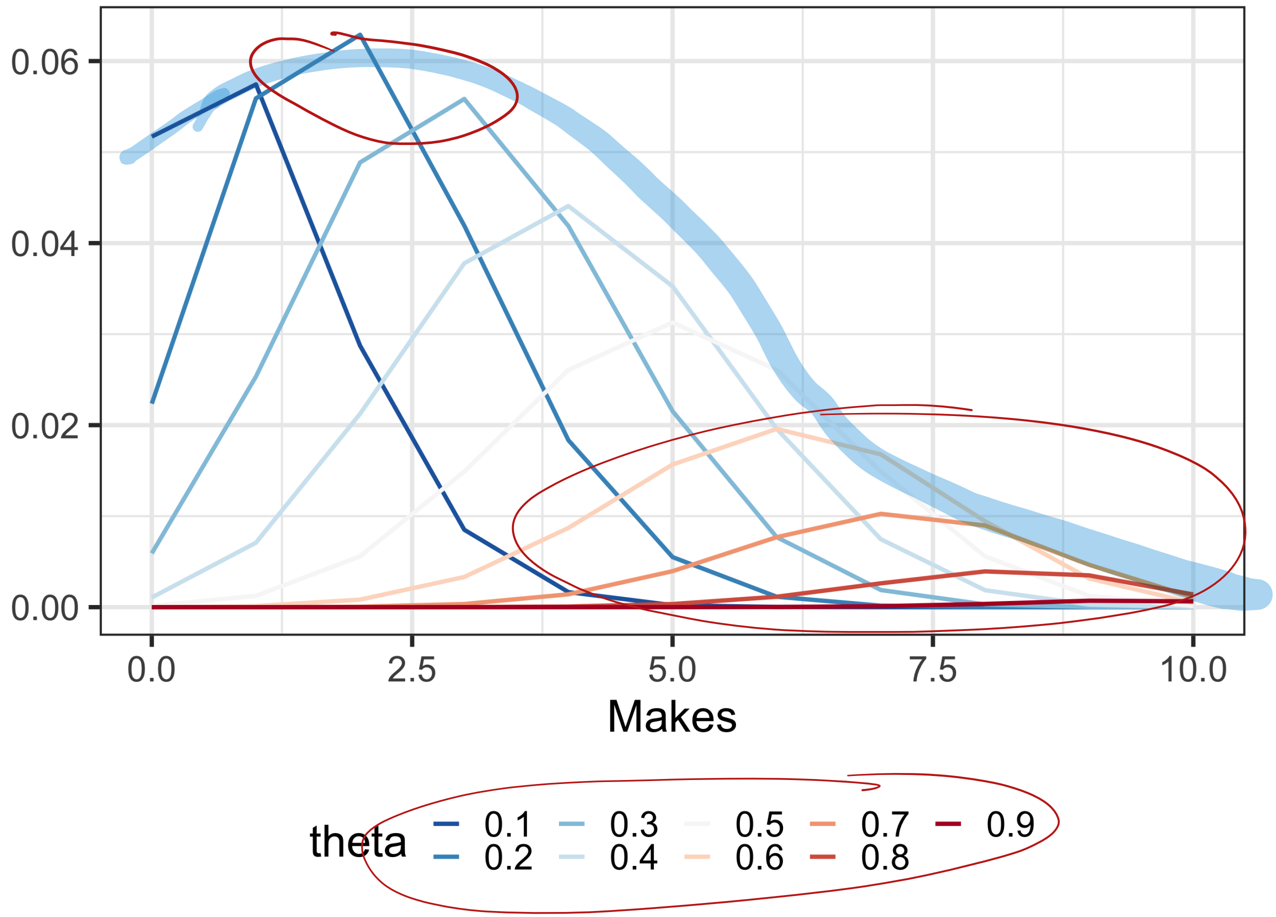
Mean $\frac{1}{1+3} = 25\%$

$\rightarrow$ Mean $\frac{2}{2+4} = 1/3$

Posterior Prediction

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

Integrate out θ.



Posterior predictive distribution

$\Gamma(n) = (n-1)!$

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

$P(\theta/y) \sim Beta(2,4)$

$$P(\tilde{y}/y) = \int_0^1 \underbrace{\binom{10}{\tilde{y}} \theta^{\tilde{y}} (1-\theta)^{10-\tilde{y}}}_{Bin} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{2-1} (1-\theta)^{4-1} \, d\theta$$

$$= \binom{10}{\tilde{y}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \underbrace{\int_0^1 \theta^{\tilde{y}+1} (1-\theta)^{13-\tilde{y}} \, d\theta}_{\text{looks like } Beta(\tilde{y}+2,\, 14-\tilde{y})}$$

$=$

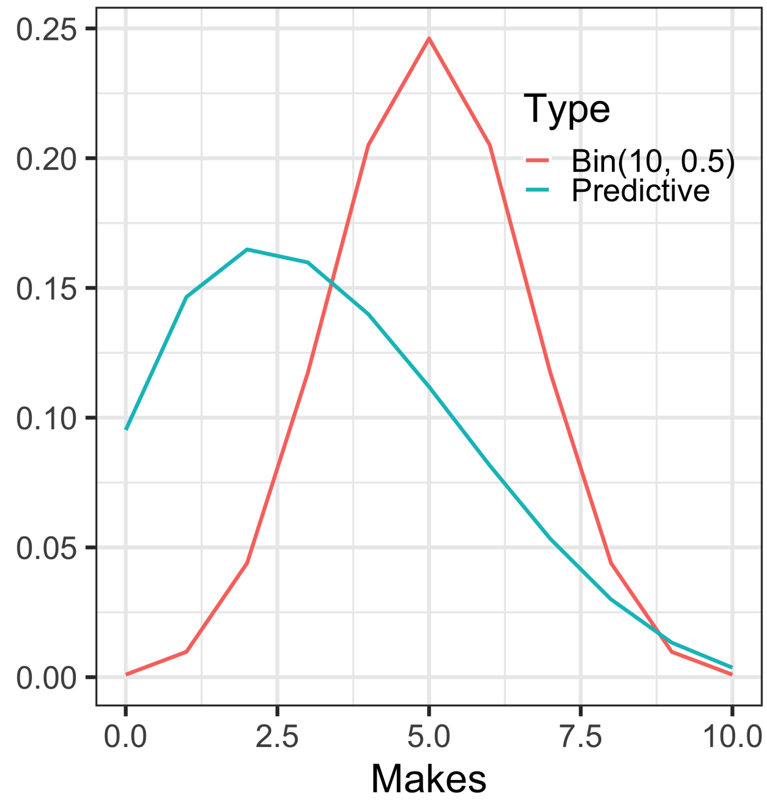
$$\int \theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$\binom{10}{\tilde{y}} \frac{\Gamma(6)}{\Gamma(2)\Gamma(4)} \frac{\Gamma(\tilde{y}+2)\Gamma(14-\tilde{y})}{\Gamma(16)}$$

Beta-Binomial Distribution

Posterior predictive distribution

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



The predictive density, $p(\tilde{y} \mid y)$, answers the question "if I take 10 more shots how many will I make, given that I already made 1 of 2".

$$\tilde{y} \sim \text{Pois}(\lambda)$$

$$y \sim \text{Pois}(\lambda)$$

$$p(\lambda) \sim \text{Gam}(a, b)$$

$$\int \underbrace{p(\tilde{y} \mid \lambda)}_{\text{pois}} \underbrace{p(\lambda \mid y)}_{\text{Gamma}} \, d\lambda$$

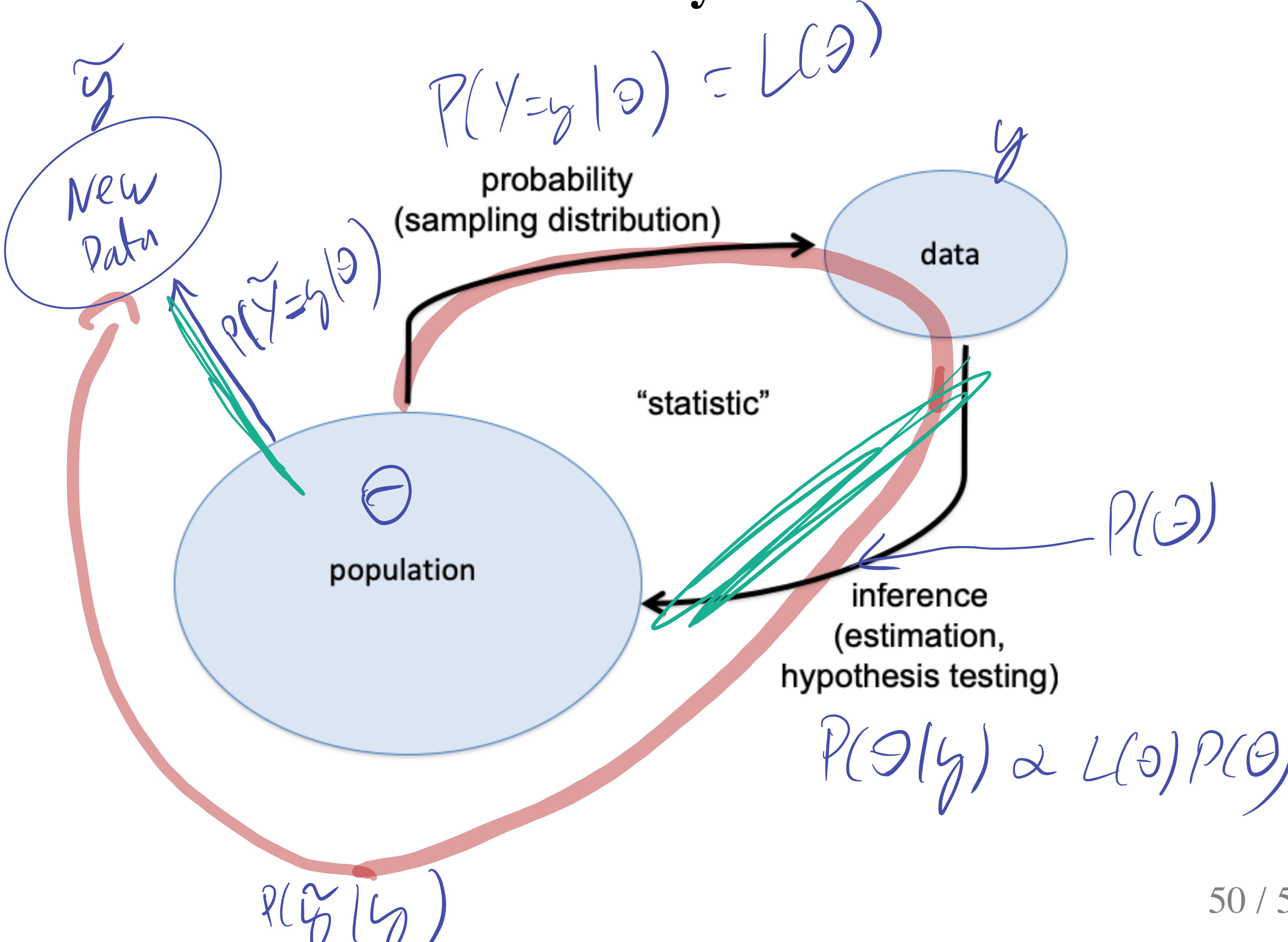
$$p(\tilde{y} \mid y) \longrightarrow \text{Negative-Bin.}$$

The posterior predictive distribution

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

- 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, $p(\theta \mid y_1, \ldots y_n)$

Posterior Predictive Density



The prior predictive distribution

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

- The prior predictive distribution describes our uncertainty about a new observation before seeing data

The prior predictive distribution

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

- The prior predictive distribution describes our uncertainty about a new observation before seeing data
- It incorporates uncertainty due to the sampling in a model $p(\tilde{y} \mid \theta)$ *and* our prior uncertainty about the data generating parameter, $p(\theta)$

Homework 1

- $\lambda \sim \text{Gamma}(\alpha, \beta)$
- $\tilde{Y} \sim \text{Pois}(\lambda)$
- $p(\tilde{y}) = \int p(\tilde{y} \mid \lambda) p(\lambda) d\lambda$ is a prior predictive distribution!
- "A Gamma-Poisson mixture is a Negative-Binomial Distribution"

Homework 1 Extra Credit

$$\begin{aligned} p(\tilde{y}) &= \int p(\tilde{y} \mid \lambda) p(\lambda) d\lambda \\ &= \int (\frac{\lambda^{\tilde{y}}}{y!} e^{-\lambda})(\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{(\alpha-1)} e^{-\beta\lambda}) d\lambda \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha) y!} \int (\lambda^{(\alpha+y-1)} e^{-(\beta+1)\lambda}) d\lambda \end{aligned}$$

$\int (\lambda^{(\alpha+y-1)} e^{-(\beta+1)\lambda}) d\lambda$ looks like an unormalized $\text{Gamma}(\alpha + y, \beta + 1)$

Summary

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